A COMPARISON OF THE WISC-R AND THE K-ABC WITH LOW-ACHIEVING ELEMENTARY SCHOOL CHILDREN

BY

JAN A. SLOAN

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN
PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF DOCTOR OF PHILOSPHY

UNIVERSITY OF FLORIDA

1984

ACKNOWLEDGEMENTS

I am sincerely grateful to a number of individuals who assisted with this project. My chairperson, Dr. Janet Larsen, has served as a positive role model throughout my graduate training at the University of Florida. She and my cochairperson, Dr. Paul Fitzgerald, consistently provided support and encouragement in moments of doubt and uncertainty. Drs. Bob Algozzine and Rus Bauer, my other committee members, helped particularly with the design of the study. In the Foundations of Education Department, Drs. Steve Olejnik and Jim Algina generously provided their time and expertise in assisting with the statistical analyses.

In the communities of Cross City and Gainesville, several individuals greatly facilitated various aspects of the study. Paula Massie and Gary Durham of the Dixie County School District provided the administrative support necessary to implement the study there, and my secretary, Cecile Howell, cheerfully photocopied needed materials. In Gainesville, Mary Ann Wagner of J.J. Finley Elementary School and Martin Mitchell of Odessey Learning Center welcomed me into their schools, and complied with my needs for student information and work space in every way possible. Their interest and efforts are very much appreciated.

One individual in particular, Steve Schonhaut, is deserving of special recognition. Besides serving as a sounding board throughout the many phases of the project, and providing assistance of a practical nature on numerous occasions, his emotional support proved invaluable. For his patience and caring I will always be deeply grateful.

TABLE OF CONTENTS

Page

ACKNOWLEDGEMENTS ii

ABSTRACT v

CHAPTER ONE INTRODUCTION 1

Need for the Study 2
Purposes of the Study 5
Rationale 6

Characteristics of the Sample 6
Characteristics of the Procedures 6
Characteristics of the Instrumentation 7
Characteristics of the Analyses 11

Definitions 12

CHAPTER TWO REVIEW OF THE RELATED LITERATURE 13

Description of the WISC-R 13

Format, Content, and Standardization 13
Theoretical Basis of Intelligence 14
Reliability 15
Validity 16
Summary 19

Description of the K-ABC 19

Format, Content, and Standardization 19
Theoretical Basis of Intelligence 22
Reliability 23
Validity 24
Summary 26

Bias in Psychological Assessment 27

Historical Perspective 27
The WISC-R and Bias 28
Summary 33
The K-ABC and Nondiscriminatory Assessment 33

CHAPTER THREE METHOD 36

Sample Selection 36
Procedures 38

Page

Research Design and Research Hypotheses 41
Methodological Assumptions 43

CHAPTER FOUR RESULTS 45

Relationships Between Intelligence Test Standard Scores and Subject Variables 45
Comparisons of Subjects' WISC-R and K-ABC Performances . . . 46

ANOVA Findings 50
Effect of Proportionality Upon the Results 53

Clinical and Educational Treatment of Test Results 59

CHAPTER FIVE DISCUSSION 63

Summary of the Findings 64
Implications and Impressions 66
Generalizability of the Findings 68
Areas for Future Investigation 69

APPENDICES

A PARENT CONSENT LETTER 72

B REVISED DUNCAN SEI SCORES FOR MAJOR OCCUPATIONAL GROUPS, 1970 CENSUS CLASSIFICATION 74

C DESCRIPTION OF THE WISC-R SUBTESTS 75

D DESCRIPTION OF THE K-ABC MENTAL PROCESSING SUBTESTS 77

E INTRODUCTION TO FIRST TESTING SESSION 78

F INTRODUCTION TO SECOND TESTING SESSION 79

G PEARSON PRODUCT MOMENT CORRELATIONS BETWEEN WISC-R AND K-ABC STANDARD SCORES 80

H GROUP MEANS, STANDARD DEVIATIONS, AND VARIANCES 81

I SUBGROUP MEANS, STANDARD DEVIATIONS, AND VARIANCES 82

J PEARSON PROUDCT MOMENT CORRELATIONS BETWEEN INTELLIGENCE TEST AND WRAT STANDARD SCORES 83

REFERENCES 84

BIOGRAPHICAL SKETCH 88

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

A COMPARISON OF THE WISC-R AND THE K-ABC WITH LOW-ACHIEVING ELEMENTARY SCHOOL CHILDREN

BY
JAN A. SLOAN

August, 1984

Chairperson: Janet J. Larsen
Cochairperson: Paul W. Fitzgerald
Major Department: Counselor Education

The Wechsler Intelligence Scale For Children-Revised (WISC-R) and the Kaufman Assessment Battery for Children (K-ABC) were compared on criteria pertinent to current issues in test bias. Forty-two elementary school children who were low-achievers in reading and/or mathematics were grouped by race (black vs. white) and social class (working vs. middle) and were individually administered the WISC-R, the K-ABC Mental Processing Scales, and the Wide Range Achievement Test (Reading and Arithmetic) in two test sessions. Analyses of the data focused on relationships between IQ and subject variables (race, socioeconomic status, achievement) for both intelligence tests, on differences between WISC-R and K-ABC performances for individual subjects by race and social class, and on the comparability of the WISC-R and the K-ABC in yielding descriptions of intellectual performance and potential placements of subjects in specific learning disability (SLD) programs.

While the tests were found to be similarly correlated with socioeconomic status and achievement, K-ABC scores bore a weaker relationship to race than WISC-R scores. Secondly, blacks averaged stronger K-ABC than WISC-R performances than whites, with differences between K-ABC and WISC-R results nonsignificant by social class. Thirdly, the tests resulted in similar descriptions of intellectual performance and similar frequencies of placed/not placed subjects in SLD programs.

The results of the study suggest that the K-ABC may be preferrable to the WISC-R with low-achieving black subjects, although the advantage offered may not change potential outcomes of intelligence testing for subjects.

CHAPTER ONE
INTRODUCTION

In the past decade the Wechsler Intelligence Scale for Children-Revised (WISC-R) (Wechsler, 1974) has frequently been the test of choice by those who use tests to evaluate the intellectual abilities of school-age children. However, although the WISC-R enjoys a strong technical backing and considerable respect among clinical and educational psychologists, its use has not been without controversy. Much criticism has been directed toward the appropriateness of this and other standardized intelligence and achievement tests with ethnic minority and disadvantaged populations. Major allegations cited against the WISC-R and other conventional tests of intelligence have included the following (Reynolds, 1982; Sattler, 1982):

(1) The content of the tests is biased in favor of white, Anglo-Saxon, middle-class children.

(2) National norms should not be used with ethnic minorities since minorities are not adequately represented in the tests' standardization samples.

(3) Language barriers and rapport problems exist between white examiners and ethnic minority subjects.

(4) Use of the tests with minorities has resulted in inequitable social consequences for them.

(5) The tests measure different constructs for ethnic minority children than for white middle-class children.

Recently, many researchers have addressed these issues, and a substantial number have concluded that on most criteria the WISC-R is not biased against ethnic minorities (Reschly, 1981; Reynolds, 1982; Sattler, 1982). Other writers, agreeing that conventional IQ tests do not contain systematic bias against blacks and others, suggest that the problem stems from areas that are not included as a part of IQ. Kaufman (1979), for example, notes that throughout the history of IQ testing, the type of items on intelligence tests has remained virtually unchanged and does not reflect important advances in cognitive psychology and neuropsychology. Sattler (1974) states, "Present and future tests are needed to evaluate the psychological processes by which children learn; tests should not be used solely to sample what has been learned and to infer the capacity to learn from their scores" (p. 45).

At a time when the WISC-R and other conventional tests of children's intelligence are viewed with considerable public and professional skepticism, a new test, the Kaufman Assessment Battery for Children (K-ABC) (Kaufman & Kaufman, 1983a,b) has emerged that promises to promote nondiscriminatory assessment. Probably the most carefully constructed major test of children's intelligence to date, the K-ABC conscientiously takes into consideration the problem of bias in assessment through separating problem-solving ability from acquired learning, de-emphasizing the role of language in test administration, including "teaching" items, and providing separate sociocultural norms. Research conducted by the authors (Kaufman & Kaufman, 1983b) with large groups of black, hispanic, and low socioeconomic status (SES) children showed that these groups earned higher average scores on the K-ABC than on the WISC-R.

Need for the Study

Probably more children are referred to educational psychologists for evaluation of learning difficulties than for any other reason. Often these children are considered "culturally disadvantaged" or "culturally deprived," as a relationship between poverty and academic failure has long been recognized (Das, 1973). Another truism is that poverty is the main characteristic of minority groups (Sattler, 1982). Consequently, in many parts of the United States, but particularly in the more economically depressed regions, children referred for psychological testing are often ethnic minorities from families of low socioeconomic status.

According to Public Law 94-142 (Education of the Handicapped Act), it is the responsibility of persons evaluating culturally different children to do so in a manner that is nondiscriminatory:

> Testing and evaluation materials and procedures used for the purposes of evaluation and placement of handicapped children must be selected and administered so as not to be racially or culturally discriminatory. (Section 121a, 530(b), USOE, 1977)

While the language appears deceptively simple, the problem is indeed complex. Public Law 94-142 does not define or provide criteria for assessment that is "nondiscriminatory."

The authors and publisher of the K-ABC claim that the test "promotes nondiscriminatory assessment"; however, it is no more clear what is meant by this claim than by the use of the concept in PL 94-142. Characteristics of the format, content, and standardization of the K-ABC suggest that it may be a less culturally loaded test than the WISC-R. Kaufman and Kaufman (1983b) also elude to this possibility upon reporting a smaller discrepancy in mean IQs between whites and blacks, and between low SES and high SES subjects for the K-ABC than for the WISC-R. However, as of yet there has been no research examining the relationships between IQ and SES and IQ and race with subjects who have taken both the K-ABC and the WISC-R. If performance on the K-ABC is less strongly related to race and SES than is performance on the WISC-R, then disadvantaged and ethnic minority children might obtain stronger K-ABC scores than WISC-R scores. As IQ is included in the definitions of exceptionalities such as educable mentally handicapped and specific learning disabled, this outcome could potentially mean that the K-ABC identifies different populations of handicapped children than the WISC-R. On the other hand, the specific changes introduced by the K-ABC may not be of sufficient significance to effect a difference in the test scores of disadvantaged and ethnic minority children. If the relationship of IQ to SES and race is approximately the same for the K-ABC and the WISC-R, there may be no particular advantage to the use of one test over the other, at least with respect to outcomes for individuals.

Besides the issue of cultural bias, psychologists evaluating children for learning deficiencies need to consider the relationship between IQ and measures of academic achievement. Public Law 94-142 stipulates that the tests used in educational assessment "are validated for the specific purpose for which they are used." Perhaps the primary justification for the use of conventional intelligence tests in school settings is that they are good predicators of school achievement. Thorndike and Hagen (1977) observe,

> The IQ does not reflect a global summation of the brain's capabilities and is certainly not an index of genetic potential, but it does predict school achievement effectively. Typical correlations between aptitude and achievement are in the 0.50-0.60 range overall, with values of about 0.70 usually obtained in elementary school. (p. 326)

In the present study the Wide Range Achievement Test (WRAT) will be used as a criterion measure of achievement. A median correlation between the WRAT and the WISC-R Full Scale IQ of about .60 has been reported for studies that included ethnic minority, and normal and mentally retarded children (Sattler, 1982). The correlation between the K-ABC Mental Processing Composite and WRAT standard scores ranges from .39 to .64 for normal subjects (Kaufman & Kaufman, 1983b). Higher correlations are evident between the K-ABC Achievement Scale and the WRAT than between the

K-ABC Mental Processing Scales and the WRAT, and Kaufman and Kaufman (1983b) state that the Achievement Scale scores are intended to be the best predictors of academic achievement in the K-ABC. However, they acknowledge that the Mental Processing Scales "also need to be shown as significant predictors of future school achievement, because any measure of present intellectual functioning should relate significantly to school success, particularly if that measure is used as one criterion for recommending specific educational interventions" (p. 120).

Thus, a question of interest and importance to psychologists evaluating children referred for learning problems might well be which test, the K-ABC or the WISC-R, is a better predictor of scholastic aptitude. If one of the intelligence tests bears a stronger relationship to academic achievement, it might be preferred over the other one, particularly if the youngster concerned is being considered for placement in a special education program.

Purposes of the Study

In this study the WISC-R and the K-ABC were compared on criteria considered relevant and important by those who evaluate children with learning difficulties. One purpose of the study was to establish relationships between IQ and SES, IQ and race, and IQ and achievement for both intelligence tests with a population of low-achievers. A second purpose was to determine if children grouped by race (black vs. white) and social class (working vs. middle) who are administered both tests perform similarly on them. The claim that the K-ABC promotes nondiscriminatory assessment was thus examined in relation to obtained test scores for individual subjects. A final purpose of the study was to compare WISC-R and K-ABC results in the larger context of potential outcomes for subjects. The questions explored were whether the two tests yield similar descriptions of intellectual performance and result in similar numbers of potential SLD placements.

Rationale

Characteristics of the Sample

The subjects selected for the study resided in Dixie County, Florida (the investigator's work setting) and in Alachua County, Florida (the investigator's residence). Characteristics of the sample were intended to reflect those common to children who are actually referred for psychoeducational evaluation of learning difficulties. Thus, elementary students 6 years 0 months through 10 years 1 month of age who demonstrated low performance in reading and/or in mathematics skills served as subjects.

Characteristics of the Procedures

Socioeconomic status (SES) of subjects was measured through parental occupation. Sociologists generally agree that occupation represents the best single measure of an individual's SES (Mueller & Parcel, 1981). In addition to being very stable over time, advantages of occupational measures of SES over other possible stratification dimensions (i.e., income and educational attainment) are that occupational information is relatively easily obtained, is relatively unobtrusive, and is less likely to be reported falsely.

As SES was an important variable in the study, questions about parental occupation were asked directly to the subjects' parents on the parent consent form (Appendix A). Secondary sources were utilized only as a last measure, and the subjects themselves were not approached for the necessary information since proxy reports of parental occupation are often unreliable (Mueller & Parcel, 1981). The questions that parents were instructed to answer were simply worded and represented slight modifications of the occupational questions that appear on the U.S. Census questionnaire.

With regard to administration of the WISC-R, the K-ABC, and the WRAT, the testing occurred in two sessions on two separate days. These arrangements were believed to be necessary to avoid fatiguing subjects and to insure that subjects would not miss an unreasonable amount of instructional time on any given day that they were taking the tests. The order in which the tests were administered was varied systematically to reduce practice effects.

Characteristics of the Instrumentation

The discussion below will relate to the selection of instruments that were used to measure subject attribute variables of interest in the study. As the WISC-R and the K-ABC were the primary focus of the study, these tests are reviewed in detail in Chapter Two (Review of the Related Literature).

The Metropolitan Achievement Test and the Comprehensive Tests of Basic Skills. The subjects who participated in the study exhibited low academic achievement, defined as a percentile rank of 25 or below for the total reading category and/or for the total mathematics category, on one of two achievement batteries--the Comprehensive Tests of Basic Skills (CTBS) and the Metropolitan Achievement Test (MAT). These two tests were selected to screen children for the sample as the CTBS was administered countywide to all regular education students in Dixie County, and the MAT to all regular education students in Alachua County. Additionally, both of these tests possess sound technical properties.

The CTBS was carefully standardized on a representative national sample of 212,000 students in public and private schools. Ahman in Buros (1972) observes, "The validity and reliability determinations follow closely the 1966 APA recommendations for psychological tests" (p. 20). Careful attention was given to content validity, and correlations with the California Achievement Test generally are in the .70 to .85 range. Kudor-Richardson reliability coefficients reported for the CTBS cluster in the .80s and .90s.

The Metropolitan Achievement Test is also generally regarded as a high quality test, although it is not without some technical flaws. The most recent edition of the MAT, the 1970 edition, was carefully developed and standardized. While the test was standardized on a national sample of school children, some writers believe that low-income and low-ability students are over-represented (Wingard & Bentler in Buros, 1978). The MAT appears to have satisfactory content validity and is highly correlated with the Otis-Lennon Mental Abilities Test. The internal consistency of the test as determined from corrected split-half and KR-21 coefficients (.84 to .97, median above .90) is good. Information on alternate forms reliability and test-retest reliability is not reported in the manual.

Both the CTBS and the MAT are group administered and are available in different levels of difficulty that correspond to grade placement. The tests utilize grade norms and report percentile ranks, standard scores, and grade equivalents.

Revised Duncan Socioeconomic Index. Revised Duncan Socioeconomic Index (Duncan SEI) scores in Stevens and Featherman (1980) were used to ascertain the socioeconomic status and social class of subjects in the study. Developed by Duncan in the early 1960s, the Duncan SEI is widely used in sociology and has been validated with respect to its relationship to commonly accepted stratification dimensions (Mueller & Parcel, 1981). Duncan SEI scores were originally derived by regressing education and income of males in the 1950 labor force on the occupational prestige rankings of 45 occupational titles. Recently, the scale has been revised to reflect changes that have occurred since 1950 in the occupational classificatory scheme and in the educational and economic characteristics of the American labor force. Stevens and Featherman (1980) provide new socioeconomic indexes for 426 occupational titles that correspond to the 1970 Census classification of occupations. The particular index that was used in the present study (TSE12) reflects the influence of income and educational attainment on the predicted percentage of "good" and "excellent" ratings for the 426 occupational titles. It is the revision recommended by Stevens and Featherman (1980), "for the study of occupational processes, regardless of whether the sample of one's interest is composed of men, women, or of both genders" (p. 19).

Use of the Duncan SEI, or of one of the revised indexes, involves coding the raw occupational data according to criteria specified by the U.S. Census Bureau. The coding procedures require time and care and a certain amount of training and experience (Featherman & Hauser, 1977). Consequently, in this study the raw data were not coded. Rather, revised Duncan SEI scores for major occupational groups (Appendix B) were assigned. Mueller and Parcel (1981) state that this system can be employed, "if the researcher has access only to crudely categorized data. . . . Although there is considerable heterogeneity of scores within each category, this is superior to impressionistically identifying cases as upper, middle, and lower class" (pp. 21-22).

The social class of subjects was derived from their revised Duncan SEI ranks, using a system analogous to that developed by Blau (1981). Employing the original Duncan index (Reiss, 1961), Blau designated respondents with scores of 0-39 as "working class," and those with scores of 40 or higher as "middle class." While in the present study the numerical criteria for designating social class differ from Blau's, reflecting the changes in occupational rankings that resulted with the revision of the Duncan SEI, the occupational categories included in the working and middle class categories are the same as Blau's.

Wide Range Achievement Test. To obtain a current measure of reading and mathematics performance relative to chronological age, the Reading and Arithmetic subtests from the Wide Range Achievement Test (WRAT) (Jastak & Jastak, 1978) were administered to all subjects. Use of the WRAT entailed advantages and disadvantages. On the plus side, the WRAT manual reports excellent split-half reliability coefficients (.94 to .98) and respectable standard errors of measurement (.88 to 1.70). With respect to concurrent validity, "numerous studies have compared the WRAT with other achievement, intelligence, and ability tests. The results of the studies suggest that the WRAT has adequate concurrent validity . . . in the range of .60 with various groups of children" (Sattler, 1982, p. 260). For the purposes of the present investigation, a primary advantage of the WRAT was that the two subtests (Reading and Arithmetic) could be administered in 15 minutes or less, an important consideration given that the other tests administered (the WISC-R and K-ABC) were lengthy and time-consuming. Additionally, the WRAT conveniently provided standard scores with the same mean and standard deviation as the WISC-R and the K-ABC.

Negative features of the WRAT concern its content and standardization. While the standardization for Level I (ages 5 through 11) involved 7,800 children, no representative national sampling was conducted. Rather, children from five eastern and two western states were tested. There has been some research (Sattler, 1982) showing that WRAT norms tend to underestimate achievement levels. With respect to content, a major weakness of the WRAT is that it provides no test of reading comprehension. In spite of these limitations, the use of the WRAT in this study was believed to be justified by its popularity as a clinical instrument and by its convenience.

Characteristics of the Analyses

One of the analyses in the study examined potential outcomes for individuals based on their K-ABC and WISC-R standard scores, presuming that all other relevant variables were held constant. The number of subjects identified as "specific learning disabled" (SLD) was compared for the WISC-R results and for the K-ABC results.

Specific learning disability in the present study was defined in accordiance with the current federal definition. According to PL 94-142 a specific learning disability,

> . . . means a disorder in one or more of the basic psychological processes involved in understanding or in using language, spoken or written, which may manifest itself in an imperfect ability to listen, think, speak, read, write, spell, or to do mathematical calculations. The term includes such conditions as perceptual handicaps, brain injury, minimal brain dysfunction, dyslexia, and developmental aphasia. The term does not include children who have learning problems which are primarily the result of visual, hearing, or motor handicaps, of mental retardation, of emotional disturbance, or of environmental, cultural, or economic disadvantage. (USOE, 1977, p. 65083)

A further stipulation of the federal definition is that the child must have "a severe discrepancy between achievement and intellectual ability." In the present study, a "severe discrepancy" was arbitrarily established as a difference of 15 or more points between the IQ and an achievement test (WRAT) standard score. An IQ-achievement discrepancy of 15 or more points was employed in the identification of specific learning disabilities in both of the school districts in which the study was conducted, and is commonly employed in others as well.

Definitions

Following are a list of terms and their definitions as used in the text of this paper:

Cultural deprivation (disadvantagement): "Substandard living conditions and/or discrimination which prevent certain individuals from participation in the cultural achievements of their society" (Wolman, 1973, p. 85).

Cultural loading: "The degree of culture specificity present in the test or individual items of the test" (Reynolds, 1982, p. 186).

IQ: "Intelligence quotient (IQ) obtained by converting raw scores on an intelligence test to a score distribution having a mean of 100 and a standard deviation, such as 16 for the Stanford-Binet or 15 for the Wechsler tests" (Sattler, 1982, p. 634). As used in this thesis, the term "IQ" refers to the standard score obtained on an intelligence test scale, and is applied ubiquitously to the WISC-R and the K-ABC even though K-ABC scores are not technically called IQs.

Minority group children: "Individuals whose values, customs, patterns of thought, language, or even interests are significantly different from the prevailing pattern of the society in which they live" (Sattler, 1974, p. 25).

Socioeconomic status (SES):

> The term "social straficiation" . . . is used to describe a social system (usually a society or community) in which individuals, families, or groups are ranked on certain hierarchies or dimensions according to their access or control over valued commodities such as wealth, power, and status. A case's relative position (and associated score) on a particular hierarchy (or combination of hierarchies) may be referred to as its SES. (Mueller & Parcel, 1981, p. 14)

Test bias: "Constant or systematic error as opposed to random error" (Reynolds, 1982, p. 178).

CHAPTER TWO
REVIEW OF THE RELATED LITERATURE

Description of the WISC-R

An important step in evaluating children referred for learning difficulties is to insure that the tests selected are appropriate. As applied in this context, "appropriate" refers to practical considerations in test administration, to the technical soundness of a test, and to its range of use. The ideal test, for example, would be racially and culturally nondiscriminatory, would be carefully normed and standardized, would possess high validity and reliability, would be easily and quickly administered and scored, and would appeal to both examiner and examinee.

Keeping in mind these standards of test perfection, in the first part of this chapter the WISC-R and the K-ABC will be introduced and described. In each case, the format and content, theoretical basis, and technical properties (standardization, reliability, validity) of the test will be discussed.

The second part of the chapter will review the problem of bias in pyschological assessment with a particular focus on the WISC-R. The K-ABCs current standing with respect to nondiscriminatory assessment will also be examined.

Format, Content, and Standardization

Originally published in 1949 as a downward extension of the adult intelligence test, the Wechsler Bellevue Intelligence Scale, the Wechsler Intelligence Scale for Children (WISC) was revised in 1974. The revision--the WISC-R--was standardized on 2200 children selected to be representative of the population on the basis of data obtained in the 1970 U.S. Census. The standardization sample ranged in age from 6½ to 16½ years, and was composed of eleven age groups, each containing 100 boys and 100 girls. Nonwhites were included in approximately the same proportions as in the U.S. population. The sample was also stratified on the variables of geographic region, occupational group of head of household, and urban versus rural residence.

The WISC-R is individually administered to children from 6 to 17 years of age. The test consists of a Verbal Scale and a Performance Scale, each comprised of five "required" subtests and one optional subtest. Descriptions of the WISC-R subtests are provided in Appendix C. Administration of the WISC-R requires from 75 to 90 minutes and yields three deviation IQs: a Verbal Scale IQ (VIQ), a Performance Scale IQ (PIQ), and a Full-Scale IQ (FSIQ), each with a mean of 100 and a standard deviation of 15. Additionally, scaled scores are available for the individual subtests.

Theoretical Basis of Intelligence

Wechsler viewed intelligence as multifaceted, and as both genetically and environmentally determined. Operationally defined, intelligence "is the aggregate or global capacity of the individual to act purposefully, to think rationally, and to deal effectively with his environment" (Wechsler, 1958, p. 7). With respect to what intelligence tests such as the WISC-R measure, Wechsler stressed the adaptability of the individual to his or her environment:

> What we measure with tests is not what tests measure--not information, not spatial perception, not reasoning ability. These are only means to an end. What intelligence tests measure, what we hope they measure, is something much more important: The capacity of an individual to understand the world about him and his resourcefulness to cope with its challenges. (1975, p. 139)

Many psychologists today would take exception to this statement. The belief that the WISC-R and other conventional tests of intelligence are primarily measures of acquired knowledge, "crystallized intelligence," is prevalent. Kaufman (1979), for example, suggests that what the WISC-R measures is based on the following premises: "1. The WISC-R subtests measure what the individual has learned. . . .2. The WISC-R subtests are samples of behavior and are not exhaustive. . . .3. The WISC-R assesses mental functioning under fixed experimental conditions" (pp. 11-12).

Some feel that there is so much confusion about the meaning of IQ test results that the term "IQ" should be changed. Mercer (1979) suggests "School Functioning Level," Reschly (1979) coined the term "Academic Aptitude," and Kaufman and Kaufman (1983a,b) use "Mental Processing Composite" as replacements for "IQ."

Another difficulty with Wechsler's theory of intelligence is that it fails to differentiate between adult intelligence and child intelligence, an omission that has been discussed as a weak point of the WISC-R (Lutey and Copeland, 1982). WISC-R items do not reflect developmental changes in intelligence, but, rather, are essentially easier versions of the same types of items that appear on the Wechsler Adult Intelligence Scale. Thus, differences between the intellectual capabilities of children and adults are viewed as quantitative rather than as qualitative.

Reliability

The reliability of the WISC-R is generally considered to be one of its strongest assets; Sattler (1982), for example, describes it as "outstanding." In the WISC-R manual, both split-half and test-retest reliability coefficients are reported. Based on the eleven age groups in the standardization sample, the average split-half reliability coefficients are .96 for the Full Scale, .94 for the Verbal Scale, and .90 for the Performance Scale. Stability coefficients were computed for a sample size of 303 children and a test-retest interval of one month: .95 for the Full Scale, .93 for the Verbal Scale, and .90 for the Performance Scale.

The manual also provides standard errors of measurement based on the entire standardization sample. The average standard errors of measurement are reported as 3.19 for the Full Scale, 3.60 for the Verbal Scale, and 4.66 for the Performance Scale.

Validity

In this section the concurrent and construct validity of the WISC-R will be reviewed. Content validity of the WISC-R (and of the K-ABC) will be discussed at a later point in relation to test bias.

Concurrent validity. There have been numerous investigations of the concurrent validity of the WISC-R; the reader is referred to Sattler (1982) for a complete listing. Regarding the results of these studies, Sattler (1982) suggests, "that when intelligence tests, receptive vocabulary tests, achievement tests, and school grades are used as criteria, the WISC-R has satisfactory concurrent validity" (p. 149). For example, Sattler reports a median correlation of .82 between the WISC-R Full Scale IQ and the Stanford-Binet IQ. The median correlation between the WISC-R Full Scale IQ and group intelligence tests (i.e., Lorge-Thorndike, Otis-Lennon, Pintner Cunningham, Revised Beta, Culture Fair Intelligence Test) is listed in Sattler (1982) as .66.

In the present study WISC-R results will be correlated with those from an achievement test, the Wide Range Achievement Test (WRAT). The results of five studies which investigated the relationships between

WISC-R scores and WRAT scores are summarized in Table 2.1. Identifying characteristics of the samples in these studies are also described.

Inspection of the data in Table 2.1 suggests moderate to strong correlations between the WISC-R Verbal IQ and WRAT standard scores for the three WRAT subtests. Likewise, a positive relationship between the WISC-R Full Scale IQ and WRAT Reading, Spelling, and Arithmetic scores is clearly evident. In most of the studies the WISC-R Performance IQ appears to be significantly related to WRAT Arithmetic scores, but bears a weaker relationship to WRAT Reading and Spelling scores.

Construct validity. The construct validity of the WISC-R is supported by the results of factor analyses of the test. Kaufman, who conducted a factor analysis of the total standardization sample (1975), summarizes the findings from his own and others' investigations:

> Regardless of the factor-analytic technique employed, the age and ethnic background of the children tested, or the nature of the sample (normal versus exceptional), factor analysis of the WISC-R has yielded one consistent and recurrent finding: the emergence of robust verbal comprehension and perceptual organization factors. The verbal factor includes significant loadings by the five regularly administered verbal subtests (although arithmetic is often a distant fifth), and the nonverbal dimension has significant loadings by all performance subtests except Coding. (Kaufman, 1982, p. 157)

Kaufman (1982) concludes that the emergence of robust factors that closely resemble Wechsler's verbal and performance scales, "delivers a vital message to clinicians: respect Wechsler's dichotomy and give much credence to the V and P IQs" (p. 158).

Table 2.1 Correlations between WISC-R Scores and WRAT Scores

WISC-R	Wide Range Achievement Test		
	Reading	Spelling	Arithmetic
Verbal IQ	1. .68	1. .62	1. .59
	2. .54	2. .49	2. .64
	3. .42	3. .21	3. .37
	4. .61	4. .65	4. .74
	5. .68	5. .61	5. .69
Performance IQ	1. .35	1. .26	1. .32
	2. .30	2. .26	2. .44
	3. .27	3. .24	3. .48
	4. .44	4. .48	4. .66
	5. .63	5. .60	5. .69
Full Scale IQ	1. .58	1. .50	1. .50
	2. ---	2. ---	2. ---
	3. .41	3. .26	3. .47
	4. .54	4. .58	4. .74
	5. .72	5. .65	5. .75

1. Covin & Lubimiv (1976); 51 caucasian children, 144 to 177 months of age, low SES background, referred for placement in childcare facility.
2. Hale (1978); 155 children, 6-16 years old, lower to middle SES, rural, referred for psychological evaluation.
3. Hartlage & Boone (1977); 13 white, 29 black children, 9-2 to 11-9 years of age, public school Grades 4 and 5.
4. Raskin, Bloom, Klee, & Reese (1978); 50 children, 6-1 to 10-11 years of age, referred for psychological evaluation.
5. Schwarting & Schwarting (1977); 282 children, ages 6-16, mixed ethnic and SES background, urban, referred for psychological evaluation.

Summary

Among clinical and school psychologists, the WISC-R is generally considered a "tried and true" test. Carefully standardized on a large national sample, the WISC-R has proven highly reliable. Another asset of considerable importance is its moderate to strong relationship with measures of academic achievement.

While WISC-R IQs are generally regarded as reliable predictors of academic aptitude, the test is believed by many to measure a rather restricted sample of behaviors. The WISC-R is essentially a test of acquired knowledge. Although it has been shown to measure the verbal and perceptual constructs that it claims to measure, its ability to tap the actual processes involved in problem solving appears limited. It is this particular limitation of the WISC-R that has been especially addressed in the K-ABC, as will be seen in the section to follow.

Description of the K-ABC

Format, Content, and Standardization

During the Spring and Summer of 1981, the K-ABC was administered to a national standardization sample of 2,000 children, 100 at each half-year of age between 2 years 6 months and 12 years 5 months. The sample was stratified by geographic region, community size, parental education, sex, and race or ethnic group based on 1980 U.S. Census Bureau data. In addition, educational placement data were obtained from the 1978 Civil Rights Survey of Elementary and Secondary Schools to insure proportional representation of exceptional children. Finally, to develop sociocultural norms, an additional 615 children (496 blacks, 119 whites) were tested and added to the standardization cases.

The K-ABC is individually administered to children 2½ to 12½ years of age. Administration time varies according to the age of the child:

35 to 50 minutes for preschoolers, 50 to 70 minutes for 5 and 6 year-olds, and 75 to 85 minutes for children 7 years of age or older. The K-ABC was designed with ease of administration in mind. Packaged in easels, it employs a similar format from one subtest to the next, consists of items that are dichotomously scored as 1 or 0, and contains materials that are colorful and highly appealing to children. Tables for the determination of subtest scaled scores, Global Scale standard scores, standard score bands for error, national percentile ranks, sociocultural percentile ranks, age equivalent scores, and grade equivalent scores are provided in an administration and scoring manual (Kaufman & Kaufman, 1983a).

Perhaps the most unique aspect of the K-ABC is the manner in which intelligence is defined. Unlike the WISC-R which measures intelligence via acquired knowledge, the K-ABC approaches intelligence in terms of an individual's style of solving problems and processing information. It is composed of four Global Scales, each yielding standard scores with a mean of 100 and a standard deviation of 15, described as follows:

1. Sequential Processing Scale:

   > Each task in this scale presents a problem which must be solved by arranging the input in a sequential or serial order. Each idea is sensibly or temporally related to the preceding one, and the linear concept of time plays an essential role in the outcome. (K-ABC National Standardization Program, 1981, p. 3)

2. Simultaneous Processing Scale:

   > The problems presented in this scale are usually spatial or analogic in nature. This input has to be integrated, and synthesized simultaneously, to produce the appropriate solution. (K-ABC National Standardization Program, 1981 p. 4)

3. Mental Processing Composite:

   This scale is "a combination of the Sequential and Simultaneous Processing Scales, yielding a global

> estimate of intellectual functioning." (Kaufman & Kaufman, 1983a, p. 4)

4. Achievement Scale:

> Items on this scale measure knowledge of facts, language concepts, and school-related skills such as reading and arithmetic.

In addition to the four Global Scales, a special Nonverbal Scale is provided to assess the mental processing of hearing-impaired, speech- and language-disordered, and non-English-speaking children. Composed of those subtests that may be administered in pantomime and responded to gesturally, the Nonverbal Scale also yields a standard score with a mean of 100 and standard deviation of 15.

Although 16 subtests are included in the K-ABC, no more than 13 are administered to any given child. Three subtests comprise the Sequential Processing Scale, seven subtests comprise the Simultaneous Processing Scale, and the Achievement Scale contains six subtests. Descriptions of each of the subtests can be found in Appendix D.

In addition to the focus on the process of problem-solving, the content of the K-ABC differs from that of the WISC-R by de-emphasizing the role of language. To minimize verbalization by the examiner and subject, most of the processing tests employ nonverbal concrete stimuli, and only three processing tests require verbal responses from the child.

Another unique feature of administration is the inclusion of "teaching items." For each of the eleven subtests in the processing scales, the unscored sample item and the first two scored items are considered teaching items: If failed by the subject, the examiner may demonstrate and explain the correct response through virtually any possible means, and allow second trials.

Theoretical Basis of Intelligence

The K-ABC was developed specifically to incorporate theoretical advances in the fields of cognitive psychology and neuropsychology. In particular, the theories of mental processing advanced by the cerebral specialization researchers, and the neuropsychological theory of Luria and his followers (Das, Kirby, & Jarman, 1979) were influential in the conception of the K-ABC.

On the basis of observations of behavioral changes following cortical lesions in humans, Luria (1966) proposed two distinct types of information processing:

> Simultaneous integration refers to the synthesis of separate elements into groups, these groups often taking on spatial overtones. The essential nature of this sort of processing is that any portion of the result is at once surveyable without dependence upon its position in the whole. . . .
>
> Successive information processing refers to processing information in a serial order. The important distinction between this type of information processing and simultaneous processing is that in successive processing the system is not totally surveyable at any point in time. Rather, a system of cues consecutively activates the components. . . .
>
> The model assumes that the two modes of processing information are available to the individual. The selection of either or both modes depend on two conditions: (1) the individual's habitual mode of processing information as determined by social-cultural and genetic factors, and (2) the demands of the task.(Das, Kirby, & Jarman, 1979, pp. 49-50)

Das and his colleagues (Das, Kirby, & Jarman, 1979) have provided empirical support for Luria's theory of simultaneous and successive information processing. Additional evidence of two distinctly different types of information processing has come from cognitive psychology and cerebral specialization research. Inspite of differences in labels given to the two processes, cognitive psychologists and neuropsychologists are generally in agreement that "for the vast majority of individuals, the left cerebral hemisphere appears to be specialized for linguistic, propositional, serial, and analytic tasks and the right hemisphere for more nonverbal, spatial, appositional, synthetic and holistic tasks" (Reynolds, 1981, p. 109).

The Kaufmans believe that intelligence is complex, and that much of intelligent behavior reflects an integration of sequential and simultaneous processing. Therefore, the K-ABC tasks were constructed so as to be primarily sequential or simultaneous in nature, but to contain elements of the other processing style as well. It should be noted that placement of a test in the Simultaneous or Successive Processing Scale depended upon the mental manipulation of the stimuli employed, not on the way the stimuli are presented to the child or the nature of the child's response.

While receiving considerable support in the literatures of several disciplines, the existence of separate, identifiable types of information processing has also been questioned. Using a special class of stochastic systems, Townsend (1971) demonstrated how serial (sequential) and parallel (simultaneous) processes can be seen to mimic one another. According to Townsend (1971), "It seems likely that within the tolerances currently accompanying psychological experimentation and modeling that some of the results now seen as supporting parallel or serial processes may be just as adequately modeled by the other type of system" (p. 163).

Reliability

Information regarding the internal consistency and stability of the K-ABC is provided in the Interpretive Manual (Kaufman & Kaufman, 1983b). The mean split-half reliability coefficients computed on school-age children (5 -12½ years) in the standardization sample are .89 for the Sequential Processing Scale, .93 for the Simultaneous Processing Scale,

.94 for the Mental Processing Composite, and .97 for the Achievement Scale. Test-retest reliability was determined by readministering the K-ABC to 246 children 2 to 3 weeks after the original testing. Average stability coefficients for the 5-0 to 8-11 age group are stated as .82 (Sequential), .88 (Simultaneous), .88 (Composite), and .95 (Achievement). Slightly stronger mean correlation coefficients were obtained for the 9-0 to 12-5 age range: .88 (Sequential), .91 (Simultaneous), .93 (Composite), and .97 (Achievement).

Standard errors of measurement are also available in the K-ABC _Interpretive Manual_. The mean standard errors of measurement for school-age children are reported as 5.0 for the Sequential Processing Scale, 4.0 for the Simultaneous Processing Scale, 3.5 for the Mental Processing Composite, and 2.7 for the Achievement Scale.

Validity

Concurrent validity. Although the K-ABC is a new test, there have already been more than 40 investigations of relationships between K-ABC scores and other intelligence test and achievement test scores (Kaufman & Kaufman, 1983b). The _Interpretive Manual_ discusses 16 studies that compared K-ABC results with WISC-R results for samples of normal, exceptional, and culturally different children. Data for a sample of 182 normal children "revealed a correlation of .70 between K-ABC Mental Processing Composite and WISC-R Full Scale IQ, and of .76 between K-ABC Achievement and WISC-R Full Scale IQ" (Kaufman & Kaufman, 1983b, p. 111). The correlation between the Simultaneous Processing standard score and WISC-R Full Scale IQ was .68, while that between Successive Processing standard score and WISC-R Full Scale IQ was .47. For a group of 121 normal children who took both the K-ABC and the Stanford-Binet Intelligence Scale, the K-ABC Mental Processing Composite correlated .61 with the Stanford-Binet IQ. Again, a higher correlation, .78, was found between the K-ABC Achievement Scale and the Binet IQ.

Two studies discussed in the Interpretive Manual examined relationships between K-ABC scores and WRAT scores with normal children. Both of these studies utilized small samples (less than 50 children), and reported somewhat different results. However, the coefficients,

> showed K-ABC Achievement to be the highest correlate of the WRAT Reading and Spelling subtests[in one of the studies][coefficients of .45 to .66] and of all three WRAT subtests [in the other study][.68 to .73]. Mental Processing Composite correlated .39 to .64 with WRAT Reading, Arithmetic, and Spelling for the two samples, and the K-ABC Sequential, Simultaneous, and Nonverbal scales also showed moderate, positive relationships with the criteria of achievement. (Kaufman & Kaufman, 1983b, p. 127)

The interested reader is referred to the Interpretive Manual for a detailed presentation of the correlation between K-ABC scores and scores from individually administered achievement tests, group-administered achievement batteries, and cognitive ability tests. In general, the results of these investigations indicate reasonably strong correlations between the criterion measures and the K-ABC Achievement Scale, and moderate correlations between the criterion measures and the K-ABC Mental Processing Composite, Sequential Processing, and Simultaneous Processing Scales.

Construct validity. Factor analyses conducted by Kamphaus, Kaufman, and Kaufman (1982) and by Wilson, Reynolds, Chatman, and Kaufman (1983) attest to the construct validity of the K-ABC. Both of these studies involved the 2,000 children in the K-ABC standardization sample. Kamphaus et al. (1982) performed principal components analyses without iterations and with varimax rotation for each of the eleven age groups in the sample. The two factors that represent sequential and simultaneous processes emerged for every age group, the relative size of each factor remaining consistent across the age range. In conducting a confirmatory analysis of the K-ABC for the eleven standardization age groups, Wilson, Reynolds, Chatman, and Kaufman (1983) analyzed two-factor solutions for the Mental Processing subtests and three-factor solutions for all K-ABC subtests combined.

> The sequential-simultaneous dichotomy was confirmed for all age groups, and the Sequential-Simultaneous-Achievement organization of the K-ABC subtests was also confirmed for all ages, including 2½ and 3-year olds. Large, highly significant values of chi-square were obtained for all analyses, and substantial factor loadings (usually in excess of .55) were found for the subtests on each factor. (Kaufman & Kaufman, 1983b, p. 107)

Summary

The K-ABC commands a number of positive features. Like its predecessor, the WISC-R, it is carefully constructed and normed. Additionally, through separating problem-solving ability from acquired learning, deemphasizing the role of language in administration, including teaching items, and providing separate sociocultural norms, the K-ABC conscientiously takes into account the problem of bias in psychological assessment. Currently, the meaning of these attributes with regard to actual outcomes for individual children is not clear.

Besides feasibility of use with diverse populations of children, the K-ABC possesses assets of a practical nature. The game-like content of the test renders it highly appealing to children without sacrificing ease in administration. Comprehensive, the K-ABC assesses both mental processes (intelligence) and school-related achievement skills.

With regard to theory, the K-ABC is based on constructs that have received considerable empirical support. This strong theoretical foundation orients interpretation for the clinical practitioner, and provides testable hypotheses for the researcher.

Finally, the K-ABC evidences technical soundness. As presented in the manual, it has very acceptable reliability. Likewise, correlations between K-ABC standard scores and other measures of intelligence and achievement are satisfactory. However, as a predictor of academic achievement, it appears that the K-ABC Mental Processing Composite does not surpass, or perhaps equal, the WISC-R Full Scale IQ. Also, it remains to be seen how the K-ABC compares with "established" tests, notably the WISC-R and the Stanford-Binet, in clinical and educational decision-making.

Bias in Psychological Assessment

Historical Perspective

Prior to 1955, the psychological testing movement enjoyed a wide degree of public acceptance. However, by the early 1970s, appropriate assessment and appropriate educational services for economically disadvantaged minority students surfaced as important and controversial issues. Criticisms of psychological testing were directed "in part at the basic logic of measurement of human abilities and in part at the effects that the measurement procedures have had upon our society" (Laosa, 1977, p. 10). As a result of the public outcry, a number of legal requirements regarding the psychological assessment of minority students were enacted; for example:

(1) Tests should be administered and interpreted in a manner consistent with the student's primary language (Diana vs. State of California, 1970; Guadalupe vs. Tempe Elementary District, 1972).

(2) Intelligence test results may not be the exclusive or primary basis for classifying children as mentally retarded; assessment must include adaptive behavior outside of school (Guadalupe vs. Tempe Elementary District, 1972).

(3) Black students may not be placed in EMR programs based on IQ test results if racial imbalance in

the composition of such programs results (Larry P. vs. Riles, 1972).

(4) Standardized intelligence tests may not be used "for the identification of black EMR children or their placement into EMR classes." (Larry P. vs. Riles, 1979, p. 104)

While some of the changes mandated by the litigation are now accepted as "good testing practices," others are still being contested in and out of court. To make matters more complex, there is disagreement among the "authorities" as to what constitutes bias. "The defining of test bias has produced considerable, as yet unresolved, debate among measurement and assessment experts" (Reynolds, 1982, p. 185).

Within this historical context, the WISC-R has been the focus of much research by psychologists attempting to prove that the test is, or is not, biased. In the discussion to follow, the findings from these research studies will be summarized. The criteria by which bias will be defined are in accordance with the model presented by Reschly (1981).

The WISC-R and Bias

Evidence of bias in the WISC-R. Reschly discusses three criteria on which conventional intelligence tests, including the WISC-R, have been found to be biased: (1) mean differences, (2) social consequences, and (3) selection ratios. The first of these, mean differences, refers to the statistical finding of higher or lower scores on the average on standardized ability and achievement tests by different sociocultural groups (Reschly, 1981). With the WISC-R, significant mean differences in IQ have been found between Anglos and non-Anglos, and between low SES and high SES groups of children (Kaufman & Doppelt, 1976; Mercer, 1979; Reschly, 1978). For example, Kaufman and Doppelt (1976) report a difference of about 16 points between the average Full Scale IQ for whites (102.3) and the average Full Scale IQ for blacks (86.4) in the

WISC-R standardization sample. For the extreme categories of SES, differences in mean IQ scores of 17 to 21 points were found.

It is perhaps ironical that mean differences in test scores among various sociocultural groups have been implicitly accepted by the courts as "bias," yet, other authorities adamantly maintain that they are not (Jensen, 1976; Reynolds, 1982). According to this alternate position, there is no _a priori_ basis for deciding that differences in mean levels of performances between Anglos and non-Anglos do, or do not, exist. In any event, group differences in tested intelligence remain long-established, well-documented findings in the literature (Havighurst and Janke, 1944; Laosa, 1977).

A second criterion of bias in the WISC-R and other traditional intelligence tests is based on few empirical studies, but considerable anectodal and historical evidence. Social consequences bias refers to the misuse of test results to justify racial, social class, and ethnic discrimination. As demonstrated by Bersoff (1979), this is "not simply an unfortunate event in American psychology that occurred a long time ago," rather, such misuses "have occurred fairly recently" (Reschly, 1981, p. 20). Social consequences bias frequently stems from over-interpretation and misinterpretation of intelligence test scores; an example of this is the mistaken notion that IQ is fixed, unitary, and predetermined genetically. Quite legitimately, it is perhaps this form of bias that has most alarmed the critics of standardized intelligence testing. Reschly (1981) notes that rather than the technical accuracy of IQ tests, "the major concerns of Black and Chicago psychologists . . . are with how tests affect the lives of persons" (p. 20).

Selection ratio bias might be considered a subtype of social consequences bias as it, too, pertains to outcomes for individuals.

Selection ratio bias relates to the disproportionate numbers of certain minorities that are determined eligible for special education programming. Historically, this definition of bias has received much legal attention primarily with respect to the overrepresentation of economically disadvantaged, minority students in educable mentally handicapped programs.

While, numerically, minority students are overrepresented in special education programs for the mildly retarded, it is not clear whether test use per se increases or decreases this overrepresentation. Recently, there is empirical evidence that "the demographic variables of race and socioeconomic status do not, independent of other pupil characteristics, influence, or bias, psychologists' diagnostic or placement behavior," and that black and low SES children are actually less likely "to be recommended for special education class placement than their white or high socioeconomic class peers with similar cognitive, behavioral, and emotional characteristics" (Reynolds, 1982, p. 182). Thus, there may be "bias" which works in reverse to keep black and low SES children out of special education programs.

Evidence against bias in the WISC-R. A very popular criticism of the WISC-R has been that the items on it are invalid for ethnic minorities. A test is presumed to have item (content) bias if "the opportunity to learn the content of [items] is not common to all environments" (Reschly, 1981, p. 12). For example, an oft cited instance of item bias on the WISC-R is the question, "What is the thing to do if a boy (girl) much smaller than yourself starts to fight with you?" While critics of this question maintain that the correct answer for a black child is to "fight back," research on this particular item has shown that it is actually easier for blacks than for whites when the children are matched for overall intellectual skill (Reynolds, 1982). Another line of research on item bias has demonstrated that rescoring IQ tests after eliminating "biased" items does not usually lead to significantly different results (Reschly, 1981).

While subjective judgments of item content by "expert representatives" from minority cultures identify many items on conventional intelligence tests as biased, agreement among judges is low and empirical evidence argues to the contrary. Reschly (1981) concludes,

> The evidence, though certainly not definitive at this point, fails to support item bias as a significant explanation for the differences in mean scores among groups. . . .Subjective judgments of item bias are not necessarily accurate, and revision of current tests either in the direction of greater or lesser cultural loading might have the undesirable effects of simultaneously increasing or maintaining group differences and reducing validity.(p. 14)

Another common complaint against standardized intelligence tests is that the testing situation for ethnic minority children is not the same as that for white, middle-class children. Reschly (1981) refers to this problem as "atmosphere bias," defined both in terms of the "kinds of responses and nature of the effort required on the test" (p. 16), and the nature of the interaction between examiner and subject. Those who maintain that atmosphere bias lowers test performance believe that ethnic minority children do not interact freely with white examiners, that these children have difficulty communicating in standard English, and that they do not appreciate the achievement aspects of the testing situation as do white, middle-class subjects (Cole and Bruner, 1971; Hardy, Welcher, Mellits, and Kagan, 1976).

The issue of atmosphere bias has been extensively researched; reviews of this literature are available in Sattler (1974, 1982) and in Lutey and Copeland (1982). Much of the research in this area was poorly designed, and the results are inconsistent and contradictory (Reschly, 1981). Those who have studied the problem in depth generally believe that if atmosphere bias exists, the effect is small (Flaugher, 1978).

Much less equivocable, investigations into psychometric properties of the WISC-R have shown that internal characteristics of the test (e.g., internal consistency reliability, test-retest reliability, subtest correlations with V, P, FSIQ scores, intercorrelations of the V, P, FSIQ scores) remain invariant regardless of sociocultural group. Similarly, factor analytic investigations have demonstrated the same factorial structures with a variety of ethinc groups (Kaufman, 1982).

The WISC-R is also currently on firm ground in terms of its predictive validity with tests of achievement. A common charge against the WISC-R is that it underestimates the competencies of ethnic minority students. If this assertion is correct, one might expect correlations of IQ with measures of academic achievement to vary as a function of group membership. However, recent studies have supported the predictive validity of the WISC-R with groups of Anglo, black, and Hispanic children (Oakland, 1977; Reschly and Reschly, 1979). Reschly (1981) concludes,

> Overall, studies on the relationship between measures such as the WISC-R and academic performance are generally positive. Clearly, the only evidence for lower or differential validity when the criterion for academic achievement is a standardized test comes from one sample of Native American students. For other groups, Anglos, blacks, and Chicanos, the WISC-R predicted standardized achievement test performance equally well regardless of group membership. . . .The presently available evidence does indicate that the WISC-R is unbiased on the criterion of predictive validity.(pp. 18-19)

Summary

The issues involved in defining and researching bias in the WISC-R are complex, and there is currently no simple, definitive answer as to whether the test is, or is not, biased against economically disadvantaged and ethnic minority children. The authorities tend to believe that on most criteria the WISC-R and other conventional intelligence tests are not biased. However, the crucial factor appears to relate not to internal characteristics of the tests themselves, but to the external use of test results by society:

> Based on the review of the WISC-R to this point, it would appear that the fundamental problem is the outcome of test use, not the test per se. This allegation, however, is serious. Failure to understand this concern has probably contributed to the poor communication between critics and proponents of tests like the WISC-R. We (proponents) have focused on various internal and external criteria of validity while the critics have raised the broader, and clearly legitimate, question of what happens to minority children as a result of test use. (Reschly, 1981, p. 25)

Kaufman and Kaufman (1983b) acknowledge the frequent empirical findings of "little racial bias in intelligence tests," but at the same time maintain that "more sensitive and thoughtful methods of intellectual assessment are required to meet the needs of minority group children" (p. 14). With this thought in mind, the discussion will now focus on features of the K-ABC that are believed to promote nondiscriminatory assessment.

The K-ABC and Nondiscriminatory Assessment

Kaufman and Kaufman (1983b) review the measures taken in designing the K-ABC to enhance use of the test with minority group children. These measures are outlined below:

(1) In the K-ABC school-related subtests, such as those requiring reading, mathematics, or general knowledge of factual information, are excluded from the Mental Processing (intelligence) scales.

(2) Teaching items are provided for all processing subtests to insure that the child understands the nature of the task involved.

(3) Answers given in subcultural slang or a foreign language are accepted.

(4) Items for the K-ABC were selected based on

> (a) the empirical results of item bias statistics, using methods developed by Angoff and Rasch, (b) the subjective perceptions and attitudes of two black and two Hispanic educators. . . , and (c) findings from the literature regarding tasks that have repeatedly been shown to be fair cross-culturally. . .or to produce minimal black-white or Hispanic-white differences. . . .(p. 15)

(5) To "prevent contamination of problem-solving ability with level of language development or fluency in verbal expression" (p. 15), the role of language in administering the K-ABC is purposely de-emphasized. Only three processing subtests require vocalized responses from subjects.

(6) Supplementary, sociocultural norms by race and sociocultural status are provided. While these additional norms are intended to aid in interpretation of test results, they are not, however, recommended as the basis for making placement decisions.

To assess the influence of these characteristics of the K-ABC on the test performances of ethnic minority and low SES subjects, Kaufman and Kaufman (1983b) analyzed the standardization data by the variables of ethnicity and socioeconomic status (SES). SES in the standardization of the K-ABC was defined by parental education. The sample was stratified according to four levels of SES, where the lowest level had less than a high school education and the highest level was college graduates. Similar to previous research on SES-IQ relationships, Kaufman and Kaufman report strong relationships between parental educational attainment and mental processing/achievement standard scores: The average Mental

Processing Composite for the lowest SES group of school-age children was 93.4 compared with 106.0 for the highest. However, mean differences of 8 to 15 points on the K-ABC Mental Processing Scales are, nevertheless, smaller than mean differences that have been observed with other intelligence tests (e.g., 17 to 21 points on the WISC-R, Kaufman & Doppelt, 1976).

For the ethnic variable, Kaufman and Kaufman (1983b) report average K-ABC standard scores of 98.2 (Sequential Processing), 93.8 (Simultaneous Processing), and 95.0 (Mental Processing Composite) for black children compared with 101.2 (Sequential Processing), 102.3 (Simultaneous Processing, and 102.0 (Mental Processing Composite) for whites. Differences in mean performance between blacks and whites are evident on the K-ABC as on the WISC-R; however, these differences are about half the size: -7.0 for the Mental Processing Composite versus -15.9 for the Full Scale IQ.

While these findings of higher average performances hold promise for black children and low SES children as groups, it is not clear what implications the findings have for individual cases. Most of the changes in test content and format introduced by the K-ABC appear to be oriented towards reducing the influence of culture on test performance. However, previous research has demonstrated that blacks perform no better (and in some instances, less well) on supposedly culture-fair tests than on conventional intelligence tests (Reynolds, 1982). Then, too, heredity, as well as environment, is related to IQ, and no test of children's intelligence can control for genetic differences between children. Consequently, whether economically disadvantaged children and ethnic minority children as individuals will obtain stronger scores on the K-ABC than on other IQ tests remains a matter for speculation.

CHAPTER THREE
METHOD

Particular features of the K-ABC and the results of preliminary research studies with groups of low SES and black children (Kaufman & Kaufman, 1983b) suggest that the K-ABC contains less cultural bias than the WISC-R. In the present study both tests were administered to a group of low-achieving elementary school children. To determine whether the two tests might be used interchangeably with this population, criteria on which the K-ABC and WISC-R were compared included relationship to SES, race, and academic achievement. Finally, the implications of test use with reference to potential outcomes for children were examined.

Sample Selection

Students who participated in the study were selected from a population of low-achieving black and white, male and female elementary school children, chronological ages 6 years 0 months through 10 years 1 month. Both public and private school children were represented in the sample. All subjects demonstrated low scholastic achievement as indicated by performance on a standardized achievement battery administered in the Spring or Fall of 1983. Two achievement tests, the Comprehensive Tests of Basic Skills (CTBS) and the Metropolitan Achievement Test (MAT) were used to determine low achievement, defined as a national percentile rank of 25 or below in the total reading category and/or total mathematics category on either the CTBS or the MAT.

Applying the above criteria, low-achieving children 6 through 9 years of age were identified in three schools (2 public, 1 private) that had agreed to participate in the project. However, only those children whose parents/guardians provided written consent for their children's participation were subsequently included in the subject pool. Also excluded were full-time special education students, students who did not speak English as a primary language, students for whom it proved impossible to ascertain SES (via parental occupation), and students whose achievement difficulties were possibly linked to uncorrected vision and hearing impairments.

Forty-two children from three elementary schools were thus included in the subject pool, all of whom were subsequently selected for the sample with two modifications. To form four proportionately balanced groups by race (black and white) and by social class (working and middle), described further in the "Procedures" section of this chapter, an "extra" black working class subject was randomly eliminated from the sample. Another adjustment to the sample was the inclusion of an additional white middle class subject obtained from a fourth public school. With these modifications to the original subject pool, the resulting sample consisted of 42 children, 18 black and 24 white.

Of the 42 children in the sample, 13 attended a public school in a rural community (Cross City, population 2,154) in Dixie County, Florida. According to the 1980 Census, the mean income for Dixie County was $12,970, the median income was $9,631, and approximately one-fourth of the families with children 5-16 years of age earned incomes below the poverty level. The remaining 29 children in the sample were enrolled in 1 of 3 schools (2 public, 1 private) located in a metropolitan area (Gainesville, population 81,371) in Alachua County, Florida. The 1980

Census lists $16,561 as the mean income for Alachua County and $12,354 as the median income, with approximately one-sixth of the families with children 5-16 years of age earning incomes below the poverty level.

In summary, the sample contained 42 children (18 black, 24 white), 7 of whom attended a private school and 36 of whom attended 1 of 3 public schools in rural and urban areas. The 11 girls and 31 boys in the sample ranged in age from 6 years 0 months to 10 years 1 month as of the initial testing session. Enrolled in regular education, the children exhibited low achievement in reading and/or mathematics on an achievement battery.

Procedures

Permission to conduct this study was granted by the University of Florida Committee for the Protection of Human Subjects. Approval to conduct the study in Alachua County was sought from the Assistant Director of the Office of Extended Services at the University of Florida, from the Director of Research of Alachua County Public Schools, and from the principals of the public elementary schools and private school involved. Permission to carry out the study in Dixie County was obtained from the superintendent and from the one elementary school principal.

In the public school system, potential subjects were identified through inspection of CTBS and MAT test scores, and school records that provided the children's birthdates and chronological ages. Permission to view this information without prior parental consent was granted by the Superintendent of Schools in Dixie County, and by the Director of Research in Alachua County. For each of the participating schools, after a group of low-achieving students was identified, consent forms were sent home with the children to be read, signed, and returned by their parents. To insure maximal cooperation, the principal of one of the schools elected to write a letter to parents in support of the project which was stapled to the consent form. Also, in both schools the children were promised a small reward (the choice of a new pencil or a quarter) provided by the experimenter if the form was returned signed. Students were instructed to give the returned forms to their classroom teachers.

The participants in the study who attended the private school were identified by the school's director who agreed to review achievement test results for his school and send home consent letters with prospective subjects. Although a letter of support was attached to the consent letter, rewards were not provided for the compliant children in this situation. As the investigator was not permitted to view school records, any needed information about the subjects was subsequently obtained directly from their parents or from the school director.

Socioeconomic status (SES) of all subjects was determined from the information provided by parents on the informed consent form (Appendix A). This information was used to ascertain the occupation of the major wage-earner in the family by referring to the 1970 Detailed Industry and Occupation Codes, U.S. Bureau of the Census in Hauser and Featherman (1977, pp. 304-317). The corresponding revised Duncan SEI score for major occupational groups (Appendix B) was then assigned, utilizing only the higher score in instances where both parents or guardians worked. Subjects with divorced parents received the SEI score of the parent with whom they resided. A few single-parent respondents who indicated that they had never been employed were instructed to answer according to the kind of work they would and could do.

Consent forms returned signed but lacking sufficient information on parental occupation were handled in one or more of the following ways:

(1) The necessary information was obtained through a telephone conversation with the parent or other adult relation.

(2) Another letter addressed to the parent requesting the needed information was sent home with the child.

(3) The questions on occupation were asked to an employee of the school who knew the child's family.

When these measures also failed to elicit the necessary information, the child in question was dropped from the study.

Children for whom it proved possible to assign a revised Duncan SEI score were also designated by social class. Analogous to the method Blau (1981) applied with the original Duncan scale, subjects who received Duncan SEI ranks of 29 or above were labeled "middle class," while those who received ranks below 29 were labeled "working class." The middle class group included professional, managerial, sales, and clerical parental occupations; the working class group was comprised of crafts, farming, transport, operatives, service (household and non-household), and labor (farm and non-farm) parental occupations.

Race of public school subjects was available from school records, while racial information for the private school children was supplied by the school director.

All children who returned signed consent forms and were assigned SEI ranks were then grouped by race and social class. The sampling procedures described in the preceding section of this chapter resulted in a sample of 42 children divided as follows:

	Social Class	
Race	Working	Middle
White	12	12
Black	12	6

During the Fall of 1983, and January, 1984, all subjects were administered the WISC-R, the K-ABC Mental Processing subtests, and the WRAT Reading and Arithmetic subtests according to directions in the respective test manuals (Wechsler, 1974; Kaufman & Kaufman, 1983a; Jastak & Jastak, 1978). The tests were given by the same examiner, a certified and experienced school psychologist, in two sessions on two different days separated by no more than 14 days.

The WISC-R was administered in one session while the K-ABC and the WRAT were administered in the other session. The order in which the tests was administered was systematically varied so that half of the subjects in each of the four groups took the WISC-R first, and the other half, the K-ABC and WRAT first. Each testing session required approximately 1 hour and 15 minutes to 1 hour and 30 minutes. Times for the testing were arranged with the children's teachers (public school subjects) or parents (private school subjects). The public school children were tested during school hours, but the private school subjects were tested after school hours, in compliance with the school director's wishes.

All subjects were provided with a standard, brief introduction to the testing session (Appendices E and F). To minimize experimenter bias, for each subject the tests were scored only after the last one in the series had been administered. All data were collected and recorded by the examiner.

Research Design and Research Hypotheses

This study was essentially descriptive and exploratory in nature. The relationships between specific subject characteristics (SES, race, academic achievement) and IQ were established for the WISC-R and for the K-ABC. The primary questions of interest were then whether correlations of the various subject variables and IQ were of similar magnitude for the two instruments. Additional analyses compared mean differences between K-ABC scores and WISC-R scores for working and middle class children, and for whites and blacks. The hypotheses tested were the following:

(1) The correlation between SES and WISC-R Full Scale IQ (FSIQ) is no different than the correlation between SES and K-ABC Mental Processing Composite (MPC).

(2) The correlation between race and FSIQ is no different than the correlation between race and MPC.

(3) The correlation between WRAT standard scores (Reading and Arithmetic) and FSIQ is no different than the correlation between WRAT standard scores (Reading and Arithmetic) and MPC.

(4) The difference between MPC and FSIQ is no different for working class subjects than for middle class subjects.

(5) The difference between MPC and FSIQ is no different for blacks than for whites.

Final analyses of particular interest and relevance to educators addressed the comparability of the WISC-R and the K-ABC in describing levels of intellectual functioning, and in identifying potential candidates for specific learning disability (SLD) programs. The hypotheses examined here were as follows:

(6) Between the WISC-R FSIQ and the K-ABC MPC, there will be no differences in the number of children described as functioning in each of the following intellectual ranges [where Wechsler's (1974) descriptor is denoted to the left of the slash, and Kaufman and Kaufman's (1983b) to the right]:

130+ (Very Superior/Upper Extreme)
120-129 (Superior/Well Above Average)
110-119 (High Average/Above Average)
90-109 (Average/Average)
80-89 (Low Average/Below Average)
70-79 (Borderline/Well Below Average)
69 and below (Mentally Deficient/Lower Extreme)

(7) Given that the criterion for a SLD diagnosis is an IQ-achievement discrepancy of at least 15 points, and all other factors are held constant, there will be no difference between the number of children classified SLD using the WISC-R and WRAT standard scores and the number classified SLD using K-ABC and WRAT standard scores.

Methodological Assumptions

This study was based on several assumptions pertaining to the sampling procedures and the population represented, and the procedures that were used with subjects. Regarding the sample, it was assumed that IQ is a function of SES background but is largely unrelated to sex, private versus public school attendance, or to urban versus rural residency. Little relationship between IQ and residency was found with the WISC-R standardization sample (Kaufman & Doppelt, 1976) and was assumed also to exist for the K-ABC. Likewise, the relationship between IQ and sex has been reported as negligible for both tests (Kaufman & Doppelt, 1976; Kaufman & Kaufman, 1983b). Another assumption was that correlations between IQ and SES, race, and achievement remain constant across the age range represented by the sample. It was assumed that the processes in operation that result in these relationships at age 6 would not change significantly in 3 to 4 years. Finally, it was assumed that the instruments used to identify low-achieving subjects were properly administered and scored.

With respect to the procedures used in the study, the assumption was made that the questions on parental occupation were answered correctly and honestly, and that revised Duncan SEI scores were assigned accurately. Furthermore, it was assumed that the division of subjects into working and middle classes reflected a true dichotomy of socioeconomic status.

Besides the methodological assumptions described above, the study entailed some inevitable methodological limitations. These are discussed in Chapter Five in relation to generalizability of the findings.

CHAPTER FOUR
RESULTS

In the present study, 42 children identified as low-achievers and grouped according to race and social class were individually administered the WISC-R, the K-ABC Mental Processing subtests, and the WRAT Reading and Arithmetic tests. Using the Statistical Analysis System (SAS) computer program, statistical tests were then carried out to compare the WISC-R and the K-ABC on criteria pertinent to current issues in test bias.

Presented and discussed in this chapter will be the results of the following analyses: relationships between IQ and important subject variables (race, SES, achievement) for both tests; race and social class comparisons of individual subjects' WISC-R and K-ABC performances; the comparability of the two instruments with respect to potential treatment of test scores. Correlations between WISC-R and K-ABC standard scores, the descriptive statistics (means, standard deviations, and variances) obtained for the sample by group and subgroup, and correlations between intelligence test standard scores and WRAT standard scores may be of interest to the reader, but were not directly relevant to the hypotheses tested. Consequently, this information is provided in Appendices G, H, I, and J, but will not be discussed.

Relationships Between Intelligence Test Standard Scores and Subject Variables

For both the WISC-R and the K-ABC, relationships between IQ and SES (revised Duncan SEI scores), race, and achievement (WRAT Reading and Arithmetic standard scores) were investigated. The correlational procedures used and the coefficients derived are presented in Table 4.1.

Inspection of Table 4.1 suggests moderate correlations between FSIQ and SES, race, and achievement. Similarly, relationships of moderate magnitude are indicated for MPC and SES, and MPC and achievement, while MPC appears to be weakly related to race. To determine if the correlation coefficients for IQ and SES, race, and achievement for the two intelligence tests are statistically similar, tests of dependent samples correlation were conducted, using the formula provided in Hinkle, Wiersma, and Jurs (1979). The results of these analyses are shown in Table 4.2.

As demonstrated by Table 4.2 only one of the analyses proved significant. The K-ABC appeared to bear a weaker relationship to race than the WISC-R, indicating that the former is less sensitive to racial influences than the latter. Thus, the hypothesis of no difference between race-FSIQ and race-MPC correlation coefficients was rejected ($p < .05$). The correlations between SES and IQ, Reading and IQ, and Arithmetic and IQ for both instruments, however, were not found to be significantly different. Among a population of low-achievers it would seem that both tests are similarly correlated with social class. Likewise, there appears to be no real difference in the abilities of the WISC-R and K-ABC to predict achievement.

Comparisons of Subjects' WISC-R and K-ABC Performances

The comparability of the WISC-R and the K-ABC was also examined in relation to differences between individual subjects' overall performances on the two tests. To determine whether subjects of particular racial and social class memberships earned higher composite scores on one or the other test, a difference score, the K-ABC Mental Processing

TABLE 4.1 Correlations between IQ and Subject Variables

Standard Score	SES[a]	Race[b]	WRAT Reading[c]	WRAT Arithmetic[c]
FSIQ (WISC-R)	.48	.47	.52	.56
MPC (K-ABC)	.43	.22	.40	.46

[a] Spearman-Rho correlation coefficient.
[b] Point-biserial correlation coefficient.
[c] Pearson product moment correlation coefficient.

TABLE 4.2 Comparison of WISC-R and K-ABC Correlation Coefficients

Relationship	WISC-R Correlation[a]	K-ABC Correlation[b]	Test Statistic[c]
SES-IQ	.48	.43	t = .52
Race-IQ	.47	.22	t = 2.635*
Reading-IQ	.52	.40	t = 1.236
Arithmetic-IQ	.56	.46	t = 1.073

[a] For the Full Scale IQ.
[b] For the Mental Processing Composite.
[c] Compared with $t_{39,.05} = 2.023$.
* $p < .05$, two-tailed.

Composite minus the WISC-R Full Scale IQ (MPC-FSIQ), was computed for all subjects. Using this difference score as the dependent variable, a 2-way analysis of variance (ANOVA) was then conducted with race (white vs. black) and social class (working vs. middle) as factors.

The 2-way ANOVA was first run with the entire sample of 42. The results, however, were suspect, as the dependent variable (MPC-FSIQ) was found to be distributed differently for one of the subgroups. Inspection of the raw data revealed that two subjects received difference scores equal to or greater than 15 points, both high negative numbers occurring in the white middle class group which inflated the variance for the group. The children in question were immature test-takers and exhibited inconsistent attending skills. One (MPC-FSIQ = -22) obtained a very high VIQ (127), an average PIQ (105), average K-ABC scores, and average to high-average WRAT scores. The second child (MPC-FSIQ = -15) received average WISC-R scores, below average K-ABC scores, and below-average WRAT scores. Examining these patterns of performance, it was concluded that the first child in question probably was not a "true" low-achiever, given his high Verbal IQ and satisfactory WRAT results, while the second child probably was. Consequently, since the homogenity of variance assumption was violated in the ANOVA of the entire sample, the suspect subject was subsequently deleted from the analysis. With this change, subgroup variances tested as not significantly different ($p > .05$), and a second analysis of variance was carried out with the sample of 41. The results of both ANOVAs ($N = 42$ and $N = 41$) will be presented, since both pro and con arguments exist in this situation for keeping/discarding extreme scores. In passing, it should be noted that the problem of extreme scores treated here had no measurable effect upon the results of the correlation or other analyses discussed in this chapter.

The question of how to interpret the performance of children who obtained stronger-than-predicted WRAT results led to a third analysis of variance. An attempt was made to differentiate "consistent" low-achievers from "inconsistent" low-achievers, the latter defined as children who "qualified" for the sample (received low CTBS/MAT scores), but who then achieved standard scores of 90 or above on the WRAT. These cases (15 of 42 children) were deleted from the analysis in this third ANOVA, leaving the children who proved the more consistent in their low achievement.

ANOVA Findings

The findings of the first ANOVA with the entire sample (N=42) are summarized in Table 4.3. As seen in Table 4.3, a significant main effect was found for race, indicating that (MPC-FSIQ) scores were not the same for whites as for blacks. The social class effect, however, was not significant; working class subjects did not obtain appreciably different MPCs and FSIQs, nor did middle class subjects. Also non-significant was the interaction between race and social class, suggesting that (MPC-FSIQ) comparisons on one of the factors did not vary across the levels of the second factor.

To pinpoint the direction of the main effect for race, least squares means for the groups were obtained, defined as "estimates of the class or subclass arithmetic means that would be expected had equal subclass numbers been obtainable" (SAS User's Guide, 1979, p. 249). These least squares means are shown in Table 4.4.

Examination of Table 4.4 indicates a larger (MPC-FSIQ) score for blacks than for whites. Moreover, blacks averaged higher MPCs across social class while whites averaged higher FSIQs.

TABLE 4.3 Two-Way ANOVA of (MPC - FSIQ) With N = 42

Source	DF	SS	F
Class	1	36.82	.64
Race	1	322.02	5.59*
Class x Race	1	54.02	.94
Error	38	2188.08	
Total	41	2654.79	

*p = .02

TABLE 4.4 Least Squares Means for (MPC - FSIQ) (N=42)

Group	Working Class	Middle Class	Row Mean
White	-3.75	.58	-1.58
Black	4.41	4.00	4.21
Row Mean	.33	2.29	

With the extreme score removed from the analysis, the 2-way ANOVA of (MPC-FSIQ) again resulted in a significant main effect for race, and a nonsignificant social class effect and class x race interaction. Least squares means and F ratios for this ANOVA appear in Table 4.5.

Comparing Tables 4.4 and 4.5, it can be seen that removal of the extreme score from the data raised the mean of the white middle class group from .58 to 2.64 while the other cell means remained unchanged. Without the high negative difference score, white middle class subjects averaged stronger MPCs, although whites as a group still averaged stronger FSIQs. Blacks had significantly larger (MPC-FSIQ) scores (favoring the MPC) than whites. While the interaction between class and race was not statistically significant ($p = .12$), blacks seemed consistent across social class in favoring the K-ABC, while working class whites tended to favor the WISC-R, and middle class whites the K-ABC.

The findings of the two 2-way ANOVAs discussed above are echoed for the third ANOVA with the sample of 27 consistent low-achievers. The least squares means and F ratios obtained are displayed in Table 4.6. Again, a significant main effect for race, nonsignificant social class effect, and nonsignificant race x social class interaction were found. Removal of the potentially higher achieving children from the sample did not alter the pattern of stronger average K-ABC - WISC-R differences for blacks administered both tests than for whites administered both tests.

Effect of Proportionality Upon the Results

The above findings that blacks, regardless of social class membership, averaged about 4 points higher on the K-ABC than on the WISC-R, while K-ABC/WISC-R differences for whites were less substantial, is

TABLE 4.5 Least Squares Means for (MPC - FSIQ) With N = 41

	Social Class		
Race	Working	Middle	Row Mean
White	-3.75^{a}	2.64^{b}	- .56
Black	4.41^{a}	4.00^{c}	4.21
Row Mean	.33	3.32	

F ratios: Race, 4.85*; Class, 1.90; Class x Race, 2.47.

[a] n = 12
[b] n = 11
[c] n = 6
[*] p = .03

TABLE 4.6 Least Squares Means for (MPC - FSIQ) for Consistent Low-Achievers (N = 27)

	Social Class		
Race	Working	Middle	Row Mean
White	-3.00[a]	.50[b]	-1.25
Black	4.56[c]	4.20[d]	4.38
Row Mean	.78	2.35	

F Ratios: Race, 4.61*; Class, .36; Race x Class, .54

[a] n = 7
[b] n = 6
[c] n = 9
[d] n = 5
* p = .04

problematic in one sense. The mean reported in Table 4.4 for blacks was obtained by taking the average of working class and middle class black means. The mean for whites was computed in the same manner. One might legitimately ask if similar means would be obtained for blacks/whites if the cell means were weighted in proportion to the number of working class and middle class blacks/whites in the population of low-achievers.

In addressing these questions, one approach might be to use the proportions observed in the present sample. However, since middle class parents are probably more inclined than working class parents to consent to their children's participation in an educational study, the present sample may contain a disproportionately large number of middle class subjects. Table 4.7 presents the actual proportional composition of the original sample (N = 42) in contrast to that of a hypothetical sample believed to be more truly representative of the population.

To determine if differences in sample composition would alter the finding of a significant main effect for race, the following exercise was carried out: For each cell in Table 4.4, the least squares mean was weighted by the hypothesized proportion of the corresponding cell in Table 4.7. Multiplying the cell means in Table 4.4 by their respective weights (the hypothesized cell proportion divided by the hypothesized row proportion), and summing across rows, resulted in the marginal means in Table 4.8.

Weighting of the cell means in Table 4.4 to adjust for the disproportional composition of the sample resulted in larger (MPC-FSIQ) differences by race than the differences actually computed. While based on hypothetical numbers, this example suggests a larger effect for race if the working class proportion of the sample is increased. The social

TABLE 4.7 Proportions of Racial and Social Class Members in Real and Hypothetical Samples

Sample 1[a]

Social Class	Race		
	White	Black	Total
Working	.285	.285	.57
Middle	.285	.142	.43
Total	.57	.43	

Sample 2[b]

Social Class	Race		
	White	Black	Total
Working	.365	.365	.73
Middle	.20	.07	.27
Total	.565	.435	

[a] Actual composition of sample (N = 42).

[b] Hypothesized composition of a representative sample.

TABLE 4.8 Weighted and Unweighted Marginal Means

Group	Unweighted Means[a]	Weighted Means[b]
White	-1.58	-2.90
Black	4.21	4.34
Working Class	.33	-1.05
Middle Class	2.29	1.47

[a] From Table 4.4.

[b] By proportional composition of a hypothetical sample (Table 4.7).

class effect, too, would appear to be stronger although would probably not reach statistical significance.

Clinical and Educational Treatment of Test Results

In clinical and educational settings, intelligence tests are administered with a specific purpose (or purposes) in mind. Frequently, the examiner seeks to ascertain the "learning apitiude" or general level of intellectual performance of a subject. Inevitably, this information is prominently presented in the psychological report, potentially affecting how those who read the report may come to regard, and behave towards, the child in question. Another common use of intelligence test scores is to establish whether a subject meets district-specific intellectual criteria for placement in special education programs.

Two of the analyses in the present study compared potential treatments of WISC-R and K-ABC results. In the first of these analyses, frequencies and percentages of subjects scoring in each of seven commonly recognized intellectual ranges were calculated for the WISC-R and for the K-ABC, as shown in Table 4.9.

While not tested for statistical significance, Table 4.9 clearly indicates that the distribution of the FSIQs and the MPCs was very comparable across the seven intellectual ranges. Based on this finding, one would not hypothesize that the two tests yield different descriptions of subjects' intellectual functioning.

Low-achieving students are often considered for special education placement, and the second of the "treatment" analyses focused on potential placement of subjects in SLD programs. SLD (specific learning disability) was operationally defined as an IQ-achievement test standard score discrepancy of at least 15, and it was assumed that other relevant factors were held constant across subjects. Using WISC-R and WRAT

TABLE 4.9 Frequencies and Percentages of Subjects Scoring in Different Intellectual Ranges (N = 42)

IQ Range	Descriptor[a]	WISC-R Distribution[b]	K-ABC Distribution[c]
130 or Above	Very Superior/Upper Extreme	0 (0%)	0 (0%)
120 - 129	Superior/Well Above Average	0 (0%)	1 (2%)
110 - 119	High Average/Above Average	2 (5%)	1 (2%)
90 - 109	Average/Average	17 (40%)	19 (45%)
80 - 89	Low Average/Below Average	15 (36%)	12 (29%)
70 - 79	Borderline/Well Below Average	7 (17%)	9 (22%)
69 and Below	Mentally Deficient/Lower Extreme	1 (2%)	0 (0%)

[a] Wechsler's (1974) descriptor is denoted to the left of the slash; Kaufman and Kaufman's (1983b) to the right.

[b] Based on Full Scale IQ.

[c] Based on Mental Processing Composite.

standard scores, and K-ABC and WRAT standard scores, the frequencies and percentages of "placed" and "not placed" subjects were then derived. This analysis was carried out in two different ways: (1) utilizing the FSIQ and MPC in all instances, (2) utilizing the FSIQ and MPC only where differences between the Verbal-Performance and Sequential-Simultaneous Scales were statistically nonsignificant ($p > .01$). Otherwise, the higher of the Verbal-Performance or Sequential-Simultaneous scores was used in the computations. These different treatments of the data were intended to reflect differences that presently exist in actual practice: Some evaluators consider the FSIQ or MPC as the best estimates of the child's intellectual aptitude, while others maintain that this is not true when a significant discrepancy is found between intelligence test scales. Taking into consideration both viewpoints, Table 4.10 presents frequencies and percentages of "placed" and not placed" subjects for the two intelligence tests.

By the first method of computation (FSIQ/MPC - WRAT standard scores), Table 4.10 shows that a large majority of subjects was not placed either by the WISC-R or K-ABC results. Furthermore, the percentages of students placed (10%) and not placed (90%) for FSIQs and MPCs were identical. A very similar pattern obtained when significant Verbal-Performance, Sequential-Simultaneous differences were taken into account: Most subjects fell into the "not placed" classification, with the number of students placed/not placed by the WISC-R and by the K-ABC differing by just one case. Thus, no apparent differences were suggested between the WISC-R and the K-ABC with respect to potential identification of learning disabilities. √

TABLE 4.10 Frequencies and Percentages of Subjects Placed SLD ($N = 42$)

Method 1[a]

WISC-R	K-ABC Not Placed	K-ABC Placed	K-ABC Total
Not Placed	37 (88.10%)	1 (2.38%)	38 (90.48%)
Placed	1 (2.38%)	3 (7.14%)	4 (9.52%)
Total	38 (90.48%)	4 (9.52%)	42 (100%)

Method 2[b]

WISC-R	K-ABC Not Placed	K-ABC Placed	K-ABC Total
Not Placed	33 (78.57%)	2 (4.76%)	35 (83.33%)
Placed	3 (7.14%)	4 (9.52%)	7 (16.67%)
Total	36 (85.11%)	6 (14.29%)	42 (100%)

[a] Placed if (FSIQ/MPC - WRAT Reading or Arithmetic) ≥ 15.

[b] Placed if (FSIQ/MPC - WRAT Reading or Arithmetic) ≥ 15, where the difference between the Verbal-Performance/Sequential-Simultaneous Scales was nonsignificant ($p > .01$). Otherwise, the higher score of the respective scales minus WRAT scores were used.

CHAPTER FIVE
DISCUSSION

In the past two decades the validity of standardized intelligence testing with ethnic minority and socioeconomically disadvantaged populations has come under close legal and professional scrutiny. Heated arguments and debates have resulted in numerous research studies focusing on various aspects of the problem. To date there are few, if any, definitive answers. The "experts" tend to claim that traditional intelligence tests like the WISC-R are not biased against minorities on most of the criteria that have been investigated. However, while the technical merits of IQ tests have been affirmed with various populations, the particularly salient question of how intelligence testing affects the lives of children has largely been ignored (Reschly, 1981).

For test examiners, perhaps the principal impact of this controversy has been to heighten sensitivity to the preferences and needs of culturally different examinees. The recent publication of the K-ABC poses interesting questions for the examiner who has long been accustomed to routine administration of the WISC-R: If, as the publisher claims, the K-ABC promotes nondiscriminatory assessment, can the same results be expected with the K-ABC as with the WISC-R? In which situations might one of the tests be preferred over the other?

The present study compared the WISC-R and the K-ABC with black and socioeconomically disadvantaged populations. The sample was comprised of the type of children most likely to come to the attention of the school psychologist--those who demonstrate low scholastic achievement.

Forty-two children who obtained low percentile ranks in reading and/or mathematics on a standardized achievement battery were grouped according to race and social class, and were individually administered the WISC-R, the K-ABC Mental Processing subtests, and the WRAT Reading and Arithmetic subtests. Analyses of the data focused on relationships between IQ and specific subject variables (SES, race, and academic achievement) for the WISC-R and for the K-ABC.

It was surmised that the test bearing the weaker relationship to race and the SES, and the stronger relationship to achievement, would be the test of choice with low-achieving students. A second analysis compared K-ABC and WISC-R results for subjects by race and social class. The questions addressed were if white working/middle class subjects and black working/middle class subjects who took both tests obtained the same, or different, scores on them. Presumably, if one test offers an advantage to a particular type of child, one would take that advantage into consideration in selecting a test to assess ability or aptitude. Finally, two of the analyses compared WISC-R/K-ABC results on potential outcomes for subjects: (1) whether the tests yield similar descriptions of overall intellectual performance, and (2) whether the tests identify the same numbers of potential SLD placements.

In the remainder of this chapter, the findings of the study will be summarized in relation to support/rejection of the hypotheses tested. Implications of the findings for the practitioner, and limitations in the generalizability of the findings will also be presented. Finally, areas for future research will be suggested and explored.

Summary of the Findings

Of the 7 hypotheses tested, 2 were rejected while 5 were not rejected, as summarized below:

(1) The hypothesis that the correlation between SES and WISC-R Full Scale IQ (FSIQ) is no different than the correlation between SES and K-ABC Mental Processing Composite (MPC) was not rejected. The correlation coefficient for FSIQ and SES (.48) was not significantly different than the correlation coefficient for MPC and SES (.43)($p > .05$).

(2) The hypothesis that the correlation between race and FSIQ is no different than the correlation between race and MPC was rejected ($p < .05$). MPC was found to be weakly related to race ($r = .22$) while FSIQ was found to be moderately related to race ($r = .47$).

(3) The hypothesis that the correlation between WRAT standard scores (Reading and Arithmetic) and FSIQ is no different than the correlation between WRAT standard scores (Reading and Arithmetic) and MPC was not rejected ($p > .05$). Correlations between Reading and IQ were .52 for the FSIQ and .40 for the MPC, while those between Arithmetic and IQ were .56 (FSIQ) and .46 (MPC).

(4) The hypothesis that the difference between MPC and FSIQ is no different for working class subjects than for middle class subjects was not rejected. Computed F ratios for the social class factor were .64 with the entire sample ($N = 42$), 1.90 for the sample with one extreme score removed ($N = 41$), and .36 for the subjects demonstrating low achievement on the CTBS/MAT and on the WRAT ($N = 27$). For all three analyses, test statistics proved to be less than the critical values of the statistics ($p > .05$).

In each of the three analyses, the interaction between race and social class was also nonsignificant ($p > .05$). However, a trend emerged for white working class subjects to average stronger WISC-R scores, and for white middle class, black middle class, and black working class subjects to average stronger K-ABC scores.

(5) The hypothesis that the difference between MPC and FSIQ is no different for blacks than for whites was rejected ($p < .05$). For the race factor, computed F ratios of 5.59, 4.85, and 4.38 were obtained for the analyses with $N = 42$, $N = 41$, and $N = 27$, respectively. For each of the three analyses, blacks consistently averaged stronger MPCs than FSIQs, while whites averaged stronger FSIQs than MPCs.

(6) The hypothesis that between the WISC-R FSIQ and the K-ABC MPC there will be no differences in the numbers of children described as functioning in each of seven intellectual ranges was not rejected. For the sample of 42, FSIQ and MPC were distributed comparably across the IQ ranges.

(7) Given that the criterion for a SLD diagnosis is an IQ-achievement discrepancy of at least 15 points, and all other factors are held constant, the hypothesis that there will be no difference between the number of children classified SLD using WISC-R and WRAT standard scores and the number classified SLD using K-ABC and WRAT standard scores was not rejected. Frequencies and percentages of children "placed" and "not placed" in SLD programs were very similar for the WISC-R and K-ABC results.

Implications and Impressions

This study was oriented towards the clinical/educational practitioner faced with the decision of which test--WISC-R or K-ABC--to administer to referred subjects. Due to the limitations of the study, described in the following section of this chapter, any conclusions drawn from the findings must be considered tentative. With this qualification in mind, the present findings suggest that the examiner would do well to pay attention to the race of low-achieving referees. It appears that black children may be offered a small advantage with the K-ABC over the WISC-R. Therefore, the examiner who must for legal reasons be careful to select tests that are nondiscriminatory, and/or who hopes to obtain a maximum estimate of aptitude, would seemingly choose the K-ABC as opposed to the WISC-R when testing a black subject referred for achievement problems. However, while this student might score a few additional points on the K-ABC than on the WISC-R, the present findings indicate that this gain would probably not alter subsequent educational/clinical treatment of the student, at least with respect to description of his intellectual functioning and eligibility for SLD programs.

The picture that emerges for white test subjects is less clear from the findings than that for blacks. In this study, low-achieving whites as a group averaged very slightly stronger WISC-R performances than K-ABC performances. With a larger sample size, however, a significant interaction between race and social class may be obtained. The data in the current study indicated a trend for working class whites to earn stronger WISC-R over K-ABC scores than their middle class white, middle class black, and working class black counterparts. If future research statistically confirms this pattern, the WISC-R may in some instances be a "better" test for working class whites than the K-ABC.

In the administration of 42 WISC-Rs and 42 K-ABCs, certain impressions were formed which will be briefly mentioned for speculation. On the part of the subjects, no clear preference emerged for one test or the other. Typically, when asked by the examiner, "Which test did you like better?" the child indicated the test most recent in his or her memory. Subjects with clear verbal strengths seemed to give stronger WISC-R than K-ABC performances; the reverse, however, of weak verbal ability, stronger K-ABC results was not evident. Children with attention deficits appeared to perform poorly on both tests, exhibiting somewhat more difficulty with the K-ABC than with the WISC-R. Finally, immature, acting-out children also seemed to have difficulty with both tests, with no apparent advantage offered by either.

From these inplications and impressions, it is concluded that the WISC-R and the K-ABC are best thought of as supplementing each other. As the Kaufmans maintain, the K-ABC is not intended to replace the WISC-R; rather, the K-ABC provides an alternative approach to the measurement of children's intelligence. Therefore, for the practitioner, the question should not be which test, WISC-R or K-ABC, is better. Instead, one should ask himself or herself how the two instruments are complementary, and in which instances one or the other is preferred.

Generalizability of the Findings

Sampling and methodological limitations of the present study may restrict the generalizability of the findings. With regard to the sample and sampling procedures, it was necessary to use only children whose parents, and schools, consented to their participation. As a random sampling procedure was not possible, the possibility exists that the children volunteered for participation in the project differed from those not permitted to participate. Therefore, the representativeness of the sample remains speculative. Another limitation concerned the necessity of using a relatively small sample size. Because of this problem, significant differences in the data may not have been detected when these differences really do exist in the population, and correlation coefficients may be attenuated due to a restricted range of scores. Consequently, the present findings are more tentative than would be the case with a large sample. A third potential problem with the sample was the variability in age. One cannot be certain that the younger children responded to the testing procedures in the same manner as the older children.

Methodological limitations of the study pertain to some of the procedures and instruments employed. While each subject's tests were scored only after the last had been administered, all of the testing was carried out by the principal investigator, and the question of experimenter bias will perhaps be raised by some. Additionally, it was not feasible to control the testing atmosphere; subjects were tested in different locations at different times of the day.

With respect to limitations of the instrumentation, the selection of the Wide Range Achievement Test as a valid and reliable measure of reading and mathematical ability might be considered questionable due to its mediocre standardization and failure to provide a measure of reading comprehension. However, the WRAT is widely used in spite of its weaknesses, and, importantly for the purposes of this study, could be quickly and easily administered.

The Duncan SEI might be considered another problematic instrument in the study. While stable, the use of occupation to measure SES assumes that the respondent does in fact have an occupation, which did not prove true in all cases. Also, the measure of SES selected (revised Duncan SEI scores for major occupational groups) lacked sensitivity to ranges of SES within occupational categories, and only the SES of one parent (the parent with the higher rank) was used. Then, too, some potential subjects in the study were lost due to parents' inability or unwillingness to read/answer the questions about occupation. Finally, rather than using occupation as the sole indicator of SES, the preferred practice would have been to obtain information on educational attainment and income in addition to occupation (Mueller & Parcel, 1981).

Areas for Future Investigation

From the preceding discussion, replication of the current findings with larger samples would seem to be a priority if firm conclusions are to be drawn. Ideally, the project described here should be undertaken on a large scale, such as an entire school district comprised of a rich ethnic and socioeconomic sampling of children.

More than providing definitive answers, the findings of the present study perhaps serve to raise a number of interesting and important questions. Given that low-achieving black youngsters do in fact perform differentially on the WISC-R and on the K-ABC, it is not at all clear why this would be so. Why would the K-ABC bear a weaker relationship to race than the WISC-R, and why would black subjects given both tests average stronger K-ABC than WISC-R performances, irrespective of social class? Recognition of the differences in content and format between the two tests, and Kaufman and Kaufman's (1983b) preliminary research findings with the K-ABC, suggest several possibilities. The Kaufmans note that the blacks in the K-ABC standardization sample earned higher Sequential Processing than Simultaneous Processing standard scores by approximately 4 points. The WISC-R contains just one subtest, Digit Span, that would be considered a "pure" measure of sequential processing, and administration of that subtest is optional. Therefore, one might deduce that if blacks as a race display a relative strength in sequential processing, they could demonstrate that strength more easily on the K-ABC than on the WISC-R. For the sample of low-achievers in the present study, however, this was not found to be true. An analysis of variance of the difference between Sequential Processing and Simultaneous Processing (SEQ-SIM, N = 41) was clearly nonsignificant for both race and social class factors.

Another conceivable explanation for the finding of higher black K-ABC scores might be that the comprehension and use of standard English is purposely de-emphasized in administering the K-ABC. Previous research with the WISC-R has indicated that language bias is not a significant factor in accounting for lower black than white performances. However, with the advent of the K-ABC (a test that makes a particularly strong effort to circumvent language barriers), the role of language in intellectual assessment probably deserves a fresh look.

Thirdly, future research may need to examine subtest profiles to explain the higher K-ABC results for blacks. Perhaps blacks as a group consistently demonstrate strengths in the abilities tapped by one or more of the K-ABC subtests.

Besides future exploration of differences in WISC-R/K-ABC performances by race and social class, a number of other research ideas surfaced while conducting this study. It became evident in testing the sample of low-achievers that the achievement difficulties of the children varied considerably in intensity and apparent causation. For some of the children, "immaturity" (poor attention to task, lack of sustained effort, excessive timidity) appeared to be the principal problem. Others showed symptoms of "true" learning disabilities, or displayed signs of emotional/behavioral disturbance. Perhaps the largest "subgroup" would be described as relatively well-behaved children who, for whatever reasons, are "slow learners." The sample, thus, was quite heterogeneous in ways other than racially and socioeconomically. Research with the subgroups of children described above might prove more meaningful and informative than that with heterogeneous samples. Other logical extensions of the present study would be studies with children actually referred for psychological testing, and studies with hispanic and other ethnic groups.

Finally, as discussed in the Review of the Literature, it is becoming more and more important to critically evaluate the "products" or results of intelligence testing, as well as the appropriateness of the tests themselves. The present study looked at the WISC-R and K-ABC in terms of two potential outcomes or treatments. There are many others which could, and should, be investigated. Applying current federal definitions of handicapping conditions, further research is needed to establish if the WISC-R and K-ABC identify the same populations of exceptional children.

APPENDIX A
PARENT CONSENT LETTER

Dear Parent,

This Fall I will be exploring two ways of measuring IQ with children from different family backgrounds. I would like to ask permission for your child, ____________________________, to participate in this research study if his or her name is selected. If you agree and your child is chosen, he or she would be given two intelligence tests plus a short reading and arithmetic test. The testing would be done on two separate days, and would take about an hour and 15 minutes on each of the days. The times for the testing would be arranged with your child's teacher, and you or your child could withdraw permission at any time. No money will be paid for participation in this project.

Your child would probably enjoy the tests. Taking the tests would also provide valuable information about your child's abilities. However, all test scores and information collected on your child would be known only to me, the researcher. If you would want the test results to be available to your child's teacher, please sign your name in the space provided on page two.

If you have any questions, or would like to know the results, a conference can be arranged with me through the school secretary.

Your cooperation and that of your child could be very helpful in better understanding different ways of measuring children's abilities. If you and your child consent for your child to participate:

(1) Please sign your name and have a witness sign his or her name.

(2) Answer the questions on occupation because the study will include information about family background.

(3) Keep this front page for your records, and return page two with your child tomorrow, if possible.

Thank you very much for your cooperation.

Jan Sloan, School Psychologist
501-E4 S.W. 75th Street
Gainesville, Fl 32607
(904) 376-2449

I have read and understand the procedure described above. I give permission for my child, ______________________________, to participate in the procedure.

____________________________ ________ ____________________ ______
(Parent) (Date) (Witness) (Date)

____________________________ ________
(Researcher) (Date)

Please answer each question below for the adults in your household who work outside of the home. If these persons are not now working, answer the questions for the type of work they usually do or have done.

(1) What kind of work does this person do? (For example: Farm manager, farm worker, mechanic, shop owner, secretary)	Child's father: ____________________
	Child's mother: ____________________
	Other relation: ____________________ (please state)
	Other relation: ____________________ (please state)
(2) What are this person's most important duties? (For example: Lays carpet, finishes concrete, hauls lumber, answers telephone)	Child's father: ____________________
	Child's mother: ____________________
	Other relation: ____________________
	Other relation: ____________________
(3) What kind of business or industry is this? (For example: Trucking, logging, farm)	Child's father: ____________________
	Child's mother: ____________________
	Other relation: ____________________
	Other relation: ____________________

Please check here _____ and sign your name below if you would like your child's test scores to be available to his or her teacher(s).

Parent signature ______________________________ Date ______________

APPENDIX B
REVISED DUNCAN SEI SCORES FOR MAJOR OCCUPATIONAL GROUPS, 1970 CENSUS CLASSIFICATION

Professional	66
Managerial	49
Sales	36
Clerical	29
Crafts	26
Farm owners and managers	24
Transport	22
Service, excl. priv. hsld	21
Laborers	19
Operatives	18
Farm laborers	17
Service, priv. hsld	15

NOTE: Based on regression equation TSE12 in *A Revised Socioeconomic Index of Occupational Status* (Center for Demography and Ecology Working Paper No. 79-48) by C. Stevens and D.L. Featherman, 1980, Madison: University of Wisconsin.

APPENDIX C
DESCRIPTION OF THE WISC-R SUBTESTS

Descriptions of the WISC-R subtests provided below are from Assessment of Children's Intelligence and Special Abilities (pp. 170-187) by J.M. Sattler, published in 1982 by Allyn & Bacon (Boston).

Verbal Scale

Information: The information subtest contains thirty questions which sample a broad range of general knowledge. . .the questions usually can be answered with a simply stated fact.

Similarities: The Similarities subtest contains seventeen pairs of words; the child must explain the similarity within each pair.

Arithmetic: The Arithmetic subtest contains eighteen orally presented problems, many of which are similar to ones often faced by children. Children are required to give their answers without using paper and pencil.

Vocabulary: In the Vocabulary subtest, which consists of thirty-two words arranged in order of increasing difficulty, the child is asked to explain orally the meaning of each word (e.g., "What is a _____?" or "What does _____ mean?").

Comprehension: The Comprehension subtest consists of seventeen questions that deal with a variety of problem situations, involving subjects such as one's body, interpersonal relations, and social mores. . . .The comprehension subtest involves comprehending given situations and providing answers to specific problems.

Digit Span: On the Digit Span, a supplementary subtest, the child listens to a series of digits given orally by the examiner and then is required to repeat the digits. There are two parts to the subtest. The Digits Forward part contains series ranging in length from three to nine digits, while the Digits Backwards part contains series ranging in length from two to eight digits.

Performance Scale

Picture Completion: The Picture Completion subtest consists of twenty-six drawings of objects from everyday life. The pictures, which are lacking a single important element, are shown one at a time. The child's task is to discover and name (or point to) the essential missing portion of the incompletely drawn picture within the twenty-second time limit.

Picture Arrangement: The Picture Arrangement subtest requires the child to place a series of pictures in a logical sequence. The twelve series (or items) are similar to short comic strips. The examiner places the individual pictures (or cards) in a specified disarranged order, and the child is asked to rearrange the pictures in the "right" order to tell a story that makes sense.

Block Design: In the Block Design subtest, the child is shown two-dimensional, red-and-white pictures of abstract designs. The task requires using blocks to assemble a design that is identical to the design on each picture. There are eleven items on the subtest. . . .

Object Assembly: The Object Assembly subtest involves the presentation of four jigsaw problems. The task consists of assembling the pieces correctly to form common objects. . . .

Coding: The Coding subtest requires the copying of symbols that are paired with other symbols. . . .The Coding subtest consists of two separate and distinct parts. Coding A is administered to children under the age of 8 years and Coding B to those who are 8 years of age and over.

Mazes: On the Mazes subtest children are required to draw a line showing how to find their way out of a series of mazes without becoming blocked (i.e., not going through a line that demarcates a wall). . . .Mazes is a supplementary test that consists of nine test mazes and one sample maze.

APPENDIX D
DESCRIPTION OF THE K-ABC MENTAL PROCESSING SUBTESTS

Descriptions of the K-ABC Mental Processing subtests provided below are from the K-ABC Sampler Manual (p. 50) by A. Kaufman and N. Kaufman, published in 1983 by American Guidance Service (Circle Pines, Minn.).

Sequential Processing Scale

Hand Movements (ages 2-6 through 12-5): Performing a series of hand movements in the same sequence as the examiner performed them.

Number Recall (ages 2-6 through 12-5): Repeating a series of digits in the same sequence as the examiner said them.

Word Order (ages 4-0 through 12-5): Touching a series of silhouettes of common objects in the same sequence as the examiner said the names of the objects. (More difficult items include an interference task between the stimulus and response.)

Simultaneous Processing Scale

Magic Window (ages 2-6 through 4-11): Identifying a picture that the examiner exposed by slowly moving it behind a narrow window, making the picture only partially visible at any one time.

Face Recognition (ages 2-6 through 4-11): Selecting from a group photograph the one or two faces that were exposed briefly on the preceding page.

Gestalt Closure (ages 2-6 through 12-5): Naming an object or scene pictured in a partially completed "inkblot" drawing.

Triangles (ages 4-0 through 12-5): Assembling several identical triangles into an abstract pattern to match a model.

Matrix Analogies (ages 5-0 through 12-5): Selecting the meaningful picture or geometric shape that best completes a visual analogy.

Spatial Memory (ages 5-0 through 12-5): Recalling the placement of pictures on a page that was exposed briefly.

Photo Series (ages 6-0 through 12-5): Placing photographs of an event in chronological order.

APPENDIX E
INTRODUCTION TO FIRST TESTING SESSION

"Hello, ____________________________, my name is Miss Sloan. I have asked your mother (father) and teacher if I could work with you today, and they said that I could. Today and on another day sometime soon, I would like to give you some tests. These tests are not like any you have taken in your classroom, and most boys and girls like them. The reason for giving the tests is to see how children from different family backgrounds do on them. You will not get a grade on the tests and they will not count on your report card, but I would like you to try your best anyway. If you get tired, thirsty, or need to use the bathroom, we can stop and rest. Is it okay with you if I give you the tests? Do you want to ask me anything before we begin?"

APPENDIX F
INTRODUCTION TO SECOND TESTING SESSION

"Hi, ______________________________. (Yesterday, several days ago, last week) you'll remember that I gave you a (some) test(s). Today I'd like to give you another (some other) test(s). Just like before, you won't get a grade for it (them), but I would like you to try your best anyway. Is it okay with you if I give you another (some other) test(s)? Do you want to ask me anything before we get started?

APPENDIX G
PEARSON PRODUCT MOMENT CORRELATIONS BETWEEN WISC-R AND K-ABC STANDARD SCORES (N=42)

	K-ABC		
WISC-R	Seq.	Sim.	MPC
VIQ	.48	.56	.61
PIQ	.49	.78	.76
FSIQ	.54	.73	.75

APPENDIX H
GROUP MEANS, STANDARD DEVIATIONS, AND VARIANCES (N=42)

	White[a]			Black[b]			Working Class[a]			Middle Class[b]		
Scale/Test	Mean	Std. Dev.	Var.	Mean	Std. Dev.	Var.	Mean	Std. Dev.	Var.	Mean	Std. Dev.	Var.
VIQ	93.04	13.80	190.56	81.22	8.05	64.77	83.54	9.07	82.35	93.89	15.20	230.93
PIQ	95.00	11.58	134.00	87.22	9.75	95.12	87.42	9.73	94.69	97.33	11.20	125.41
FSIQ	93.75	11.60	134.63	82.89	7.99	63.87	84.29	8.40	70.56	95.50	12.08	145.91
Seq.	90.92	12.38	153.30	88.06	11.26	126.76	86.04	12.41	154.04	94.56	9.32	86.85
Sim.	94.58	12.77	163.04	88.94	9.61	92.29	86.58	7.78	60.60	99.61	12.14	147.43
MPC	92.17	12.75	162.49	87.17	8.86	78.50	84.63	8.46	71.55	97.22	10.98	120.65
Reading	95.13	13.52	182.72	87.28	14.27	203.51	89.92	14.90	221.99	94.22	13.28	176.42
Arithmetic	94.79	10.19	103.83	89.61	8.61	74.13	89.21	9.01	81.13	97.06	9.16	83.82

[a] n = 24

[b] n = 18

APPENDIX I
SUBGROUP MEANS, STANDARD DEVIATIONS, AND VARIANCES (N=42)

	White						Black					
	Working Class[a]			Middle Class[a]			Working Class[a]			Middle Class[b]		
Scale/Test	Mean	Std. Dev.	Var.	Mean	Std. Dev.	Var.	Mean	Std. Dev.	Var.	Mean	Std. Dev.	Var.
VIQ	87.08	10.43	108.81	99.00	14.57	212.18	80.00	6.00	36.00	83.67	11.41	130.27
PIQ	90.25	9.70	94.02	99.75	11.70	136.93	84.58	9.30	86.45	92.50	9.12	83.10
FSIQ	87.75	8.97	80.39	99.75	11.07	122.57	80.83	6.41	41.06	87.00	9.82	96.40
Seq.	84.50	11.99	143.73	97.33	9.33	86.97	87.58	13.16	173.17	89.00	6.96	48.40
Sim.	86.92	8.25	68.08	102.25	12.02	144.57	86.25	7.64	58.39	94.33	11.54	133.07
MPC	84.00	7.86	61.82	100.33	11.51	132.42	85.25	9.32	86.93	91.00	7.01	49.20
Reading	92.00	13.40	179.45	98.25	13.46	181.30	87.83	16.59	275.24	86.17	9.17	84.17
Arithmetic	91.92	9.52	90.63	97.67	10.41	108.42	86.50	7.94	63.00	95.83	6.62	43.77

[a] n = 12

[b] n = 6

APPENDIX J
PEARSON PRODUCT MOMENT CORRELATIONS BETWEEN INTELLIGENCE TEST AND WRAT STANDARD SCORES (N=42)

	K-ABC			WISC-R		
WRAT	Seq.	Sim.	MPC	VIQ	PIQ	FSIQ
Reading	.54	.22	.40	.52	.35	.52
Arithmetic	.51	.32	.46	.51	.46	.56

REFERENCES

Benton, A.L. Developmental dyslexia: Neurological aspects. In W.J. Friedlander (Ed.), Advances in neurology (Vol. 7). New York: Raven Press, 1975.

Bersoff, D. Regarding psychologists testily: Legal regulation of psychological assessment in the public schools. Maryland Law Review, 1979, 39, 27-120.

Blau, Z.S. Black children/white children. New York: The Free Press, 1981.

Buros, O.K. (Ed.). The seventh mental measurements yearbook. Highland Park, N.J.: Gryphon Press, 1972.

Buros, O.K. (Ed.). The eighth mental measurements yearbook. Highland Park, N.J.: Gryphon Press, 1978.

Cole, M., & Bruner, J.S. Cultural differences and inferences about psychological processes. American Psychologist, 1971, 26, 867-876.

Covin, T.M., & Lubimiv, A.J. Concurrent validity of the WRAT. Perceptual and Motor Skills, 1976, 43, 573-574.

Das, J.P. Cultural deprivation and cognitive competence. In N.R. Ellis (Ed.), International review of research in mental retardation (Vol. 6). New York: Academic Press, 1973.

Das, J.P., Kirby, J., & Jarman, R.F. Simultaneous and successive cognitive processes. New York: Academic Press, 1979.

Diana vs. State Board of Education, C-70 37 RFP, District Count for Northern California (February, 1970).

Featherman, D.L., & Hauser, R.M. The measurement of occupation in social surveys. In R. Hauser & D. Featherman (Eds.), The process of stratification. New York: Academic Press, 1977.

Flaugher, R. The many definitions of test bias. American Psychologist, 1978, 33, 671-679.

Guadalupe vs. Tempe Elementary School District, 71-435, District Court for Arizona, January, 1972.

Hale, R.L. The WISC-R as a predictor of WRAT performance. Psychology in the Schools, 1978, 15, 172-175.

Hardy, J.B., Welcher, D.W., Mellits, E.D., & Kagan, J. Pitfalls in the measurement of intelligence: Are standard intelligence tests valid instruments for measuring the intellectual potential of urban children? Journal of Psychology, 1976, 94, 43-51.

Hartlage, L.C., & Boone, J.E. Achievement test correlates of Wechsler Intelligence Scale for Children and Wechsler Intelligence Scale for Children-Revised. Perceptual and Motor Skills, 1977, 45, 1283-1286.

Hauser, R., & Featherman, D. The process of stratification. New York: Academic Press, 1977.

Havighurst, R.J., & Janke, L.L. Relations between ability and social status in a midwestern community. I. Ten-year-old children. Journal of Educational Psychology, 1944, 35, 357-368.

Hinkle, D.E., Wiersma, W., & Jurs, S.G. Applied statistics for the behavioral sciences. Chicago: Rand McNally, 1979.

Jastak, J.F., & Jastak, S. The Wide Range Achievement Test (rev. ed.). Wilmington, Del.: Jastak Associates, 1978.

Jensen, A.R. Test bias and construct validity. Phi Delta Kappan, 1976, 58, 340-346.

Kamphaus, R.W., Kaufman, A.S., & Kaufman, N.L. A cross-validation study of sequential-simultaneous processing at ages 2½ - 12½ using the Kaufman Assessment Battery for Children (K-ABC). Paper presented at the 90th Annual Convention of the American Psychological Assoc., Washington, D.C., 1982.

Kaufman, A.S. Factor analysis of the WISC-R at eleven age levels between 6½ and 16½ years. Journal of Consulting and Clinical Psychology, 1975, 43, 135-147.

Kaufman, A.S. Intelligent testing with the WISC-R. New York: John Wiley & Sons, 1979.

Kaufman, A.S. The impact of WISC-R research for school psychologists. In C.R. Reynolds & T.B. Gutkin, The handbook of school psychology. New York: John Wiley & Sons, 1982.

Kaufman Assessment Battery for Children national standardization program examiner's manual. Circle Pines, Minn.: American Guidance Service, 1981.

Kaufman, A.S., & Doppelt, J.E. Analysis of WISC-R standardization data in terms of the stratification variables. Child Development, 1976, 47, 165-171.

Kaufman, A.S., & Kaufman, N.L. Kaufman Assessment Battery for Children administration and scoring manual. Circle Pines, Minn.: American Guidance Services, 1983a.

Kaufman, A.S., & Kaufman, N.L. Kaufman Assessment Battery for Children interpretive manual. Circle Pines, Minn.: American Guidance Service, 1983b.

Kaufman, A.S., & Kaufman, N.L. Kaufman Assessment Battery for Children sampler manual. Circle Pines, Minn.: American Guidance Service, 1983c.

Laosa, L.M. Nonbiased assessment of children's abilities. In T. Oakland, Psychological and educational assessment of minority children. New York: Brunner/Mazel, 1977.

Larry P. et al. vs. Wilson Riles et al. United States District Court, Northern District of California, Case No. C-71-2270 RFP, 1972, 1974, 1979.

Luria, A.R. Higher cortical functions in man. New York: Basic Books, 1966.

Lutey, C., & Copeland, E. Cognitive assessments of the school-age child. In C.R. Reynolds & T.B. Gutkin, The handbook of school psychology. New York: John Wiley & Sons, 1982.

Mercer, J. Technical manual: SOMPA: System of multicultural pluralistic assessment. New York: Psychological Corporation, 1979.

Mueller, C.W., & Parcel, T.L. Measures of socioeconomic status: Alternatives and recommendations. Child Development, 1981, 52, 13-30.

Oakland, T. (Ed.). Psychological and educational assessment of minority children. New York: Brunner/Mazel, 1977.

Raskin, L.M., Bloom, A.S., Klee, S.H., & Reese, A. The assessment of developmentally disabled children with the WISC-R, Binet, and other tests. Journal of Clinical Psychology, 1978, 34, 111-114.

Reiss, A.J., Jr. (Ed.). Occupations and social status. New York: The Free Press, 1961.

Reschly, D. WISC-R factor structures among Anglos, Blacks, Chicanos, and Native American Papagos. Journal of Consulting and Clinical Psychology, 1978, 46, 417-422.

Reschly, D. Nonbiased assessment. In G. Phye & D. Reschly (Eds.), School psychology: Perspectives and issues. New York: Academic Press, 1979.

Reschly, D. Nonbiased assessment. Paper presented at the Annual Convention of the Florida Association of School Psychologists, Daytona Beach, Florida, November 1981.

Reschly, D., & Reschly, J. Validity of WISC-R factor scores in predicting teacher ratings of achievement and attention among four groups. Journal of School Psychology, 1979, 17, 355-361.

Reynolds, C.R. The neuropsychological basis of intelligence. In G.W. Hynd & J.E. Obrzut (Eds.), Neuropsychological assessment and the school-age child. New York: Grune & Stratton, 1981.

Reynolds, C.R. The problem of bias in psychological assessment. In C.R. Reynolds & T.B. Gutkin (Eds.), The handbook of school psychology. New York: John Wiley & Sons, 1982.

SAS user's guide, 1979 edition. Raleigh, N.C.: SAS Institute Inc., 1979.

Sattler, J.M. Assessment of children's intelligence. Philadelphia: W.B. Saunders, 1974.

Sattler, J.M. Assessment of children's intelligence and special abilities (second ed.). Boston: Allyn and Bacon, 1982.

Schwarting, F.C., & Schwarting, K.R. The relationship of the WISC-R and WRAT: A study based upon a selected population. Psychology in the Schools, 1977, 14, 431-433.

Stevens, G., & Featherman, D.L. A revised socioeconomic index of occupational status (Center for Demography and Ecology Working Paper No. 79-48). Madison: University of Wisconsin, 1980.

Thorndike, R.L., & Hagen, E. Measurement and evaluation in psychology and education, 4th ed. New York: John Wiley & Sons, 1977.

Townsend, J.T. A note on the identifiability of parallel and serial processes. Perception and Psychophysics, 1971, 10, 161-163.

United States Office of Education (USOE). Assistance to states for education of handicapped children: Procedures for evaluating specific learning disabilities. Federal Register, 1977, 42, 65082-65085.

Wechsler, D. The measurement and appraisal of adult intelligence. Baltimore: Williams & Wilkins, 1958.

Wechsler, D. Wechsler Intelligence Scale for Children-Revised. New York: Psychological Corporation, 1974.

Wechsler, D. Intelligence defined and undefined: A relativistic appraisal. American Psychologist, 1975, 30, 135-139.

Wilson, V.L., Reynolds, C.R., Chatman, S., & Kaufman, A.S. Confirmatory analysis of simultaneous, sequential, and achievement factors on the K-ABC. Paper presented at the meeting of the National Association of School Psychologists, Detroit, March 1983.

Wolman, B.B. Dictionary of behavioral science. New York: Van Nostrand Reinhold, 1973.

BIOGRAPHICAL SKETCH

Born July 22, 1950, Jan Sloan was raised in Chicago, Illinois; Madison, Wisconsin; Edina, Minnesota; and Jackson, Michigan. She graduated from Parkside High School in Jackson, Michigan,in June, 1968.

Jan attended the University of Texas at Austin where she majored in psychology and graduated with honors in December, 1971. In 1975, she was awarded the M.A. degree in school psychology from Western Michigan University, Kalamazoo, Michigan. Following an internship in Naples, Florida, she moved to Gainesville where her formal education continued at the University of Florida. She received the Ed.S. degree in school psychology in 1980.

Jan is a Florida certified and licensed school psychologist and has been employed in this capacity since 1976. Her interest in nonbiased assessment stemmed from her work with culturally disadvantaged children in the Marion County, Gilchrist County, and Dixie County, Florida school districts. She is an affiliate of the Florida Association of School Psychologists and of the National Association of School Psychologists.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Janet J. Larsen, Ed.D., Chairperson
Professor of Counselor Education

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Paul Fitzgerald

Paul W. Fitzgerald, Ed.D., Cochairperson
Professor of Counselor Education

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Robert F. Algozzine

Robert F. Algozzine, Ph.D.
Professor of Special Education

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Russell M. Bauer

Russell M. Bauer, Ph.D.
Assistant Professor of Clinical
Psychology

This dissertation was submitted to the Graduate Faculty of the Department of Counselor Education in the College of Education and to the Graduate School, and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

April, 1984

Dean for Graduate Studies and Research